AFRL-RH-WP-TR-2012-0107

ANALYST PERFORMANCE MEASURES: VOLUME II: INFORMATION QUALITY TOOLS FOR PERSISTENT SURVEILLANCE DATA SETS

Marina Altynova
Ed Wasser
Telford Berkey
Dr. Sanjay Boddhu
Tin Sa
Qbase, LLC
2619 Commons Boulevard
Dayton OH 45431

Brian Tsou
Forecasting Division
Human-Analyst Augmentation Branch

OCTOBER 2011
Final Report

AIR FORCE RESEARCH LABORATORY
711™ HUMAN PERFORMANCE WING,
HUMAN EFFECTIVENESS DIRECTORATE,
WRIGHT-PATTERSON AIR FORCE BASE, OH 45433
AIR FORCE MATERIEL COMMAND
UNITED STATES AIR FORCE

NOTICE AND SIGNATURE PAGE

AFRL-RH-WP-TR-2012-0107 HAS BEEN REVIEWED AND IS APPROVED FOR PUBLICATION IN ACCORDANCE WITH ASSIGNED DISTRIBUTION STATEMENT.

//SIGNED//
BRIAN TSOU
Work Unit Manager
Human-Analyst Augmentation Branch

//SIGNED//
LOUISE A. CARTER, Ph.D
Chief, Forecasting Division
Human Effectiveness Directorate
711th Human Performance Wing
Air Force Research Laboratory

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. **REPORT DATE** *(DD-MM-YYYY)* 07-OCT-2011	2. **REPORT TYPE** Final	3. **DATES COVERED** *(From - To)* June 2009 – September 2011
4. **TITLE AND SUBTITLE** Analyst Performance Measures Volume II: Information Quality Tools for Persistent Surveillance Data Sets		**5a. CONTRACT NUMBER** FA8650-09-6939 0006 **5b. GRANT NUMBER** **5c. PROGRAM ELEMENT NUMBER** 62202F
6. **AUTHOR(S)** Marina Altynova, Ed Wasser, Telford Berkey, Dr. Sanjay Boddhu, Tin Sa, Brian Tsou		**5d. PROJECT NUMBER** 7184 **5e. TASK NUMBER** 06 **5f. WORK UNIT NUMBER** 7184X18W
7. **PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)** Qbase, LLC 2619 Commons Boulevard Dayton OH 45431		8. **PERFORMING ORGANIZATION REPORT NUMBER**
9. **SPONSORING / MONITORING AGENCY NAME(S) & ADDRESS(ES)** Air Force Materiel Command Air Force Research Laboratory 711th Human Performance Wing Human Effectiveness Directorate Forecasting Division Human-Analyst Augmentation Branch Wright-Patterson AFB OH 45433		10. **SPONSOR/MONITOR'S ACRONYM(S)** 711 HPW/RHXM 11. **SPONSOR/MONITOR'S REPORT NUMBER(S)** AFRL-RH-WP-TR-2012-0107

12. **DISTRIBUTION / AVAILABILITY STATEMENT**
Distribution A: Approved for Public Release

13. **SUPPLEMENTARY NOTES**
88ABW-2012-4361, dated 8 August 2012

14. **ABSTRACT**
The Air Force desires a comprehensive vehicle to identify and address requirements for information quality tools and techniques to support defensive and offensive operations research in the layered sensing domain. As use of remote sensors in the Air and Space domains increases, the value of sensor datasets must be maximized, and assurances established that the product outcomes meet the application requirements. As multiple sensors are combined into layered sensing systems, this increases the need to understand not only the quality and fitness for using the individual sensor data streams, but also how to assess the quality and value of the aggregate data.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF: UNCLASSIFED			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. **REPORT**	b. **ABSTRACT**	c. **THIS PAGE**	SAR	70	Brian Tsou
U	U	U			**19b. TELEPHONE NUMBER** *(include area code)* NA

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. 239.18

THIS PAGE LEFT INTENTIONALLY BLANK.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1.0 SUMMARY

As the use of remote persistent sensing technology rapidly increases in both the government and commercial sectors, an ever increasing amount of sensor data is being generated from multiple sensor layers (such as ground, air, and space) and in a variety of modalities (infrared, hyper-spectral, synthetic aperture radar, audio, and so on).

As Lt. General David Deptula, the Air Force's Deputy Chief of Staff for Intelligence, Surveillance, and Reconnaissance, recently stated, "We are going to be swimming in sensors and drowning in data."[1]

One implication of this explosion of sensor data is that there is an increasing need to rely upon sophisticated exploitation algorithms to process and filter the raw sensor data to generate information that can aid a human analyst in determining whether or not a situation exists that requires some action to be taken.

The quality of information generated by these exploitation algorithms is dependent upon the quality of data collected from the sensor as well as upon factors external to the sensor such as the day of the year, time of day, weather conditions, speed, location, and altitude of the sensor platform.

When the information generated by these algorithms is presented to the analyst, it is essential that he or she has some indication of level of quality of that information so that a fully informed decision can be made. One extreme example of what can happen when decisions are made on the basis of low quality information is the downing of Iran Air Flight 655 by the USS Vincennes in 1988.

Although a number of factors contributed to this tragedy, it's clear that the decision makers in this incident lacked a clear understanding of the quality of the information presented to them[2]. Similarly, the targeting and bombing of the Chinese embassy in Belgrade in 1999 illustrates how low quality information, "impressively packaged," gave decision makers an impression that the information was of much higher quality than it actually was.[3]

The objective of the Information Quality Tools for Persistent Surveillance Data Sets program is to investigate tools and methods for measuring, aggregating, quantifying and communicating metrics that accurately represent the level of quality (accuracy, precision, timeliness, trustworthiness, and so on) associated with the data collected by persistent sensors and the information derived from those sensors by exploitation algorithms.

During the first year of this program, our focus has been on identifying and quantifying the characteristics of sensor data that impact the quality of information provided for computer and human analysis. In addition, we have looked at means of calculating, storing, and communicating

[1] Politics | Air Force Develops New Sensor to Gather War Intel | Seattle Times Newspaper
[2] USS Vincennes Incident - Dan Craig, Dan Morales, Mike Oliver
[3] Chinese Embassy Bombing - A Wide Net of Blame - NYTimes.com

these information quality metrics along with the sensor data so that this information is preserved and made available to the analyst as part of the decision making process.

During the second year of this program, our focus expanded to include a means of calculating, storing, and communicating the Value of Information (VOI). The decision to expand our focus was driven largely by feedback from AFRL personnel during briefings we conducted early in the year to review the work performed during the first year of this program. Much of the feedback we received had to do with the tremendous volume of sensor data being collected and the limited number of analysts available to review the data. As a result, there is a need to reduce the volume and increase the value of the information being presented. Even though we may have lots of very high quality data available, if that information provides no added value to the end-user's mission, there is no need to present it to them. Indeed, presenting this information could distract them from more valuable information that is also available from other sensors or information sources.

Unlike quality of information, determination of the value of information is very much dependent upon the goals of the mission and reflects such characteristics as the usefulness, uniqueness, and relevance of the information to the mission at hand.

Increasingly algorithms are being developed that perform sophisticated analysis to determine if something of interest is occurring (e.g. establishing a baseline of normal activity and identifying significant deviations from the baseline that could represent the occurrence of abnormal activity) and to recognize certain types of patterns (e.g. shapes, sizes, movements) in order to identify people and objects that might be of interest. Typically these algorithms generate metrics that can be used to help determine the value of the information. For example a pattern matching algorithm might provide a metric that represents how certain the algorithm is that it has detected a match. This metric can then be used to help determine the value of the information for different mission objectives. For instance if an analyst is being presented with several items of interest at the same time, those matches with a high certainty metric could be prioritized over those that have low certainty values.

We also investigated how data quality, information quality, and value of information fit into a situational awareness framework with a focus on what quality factors are utilized for the different stages of establishing situational awareness (perception, comprehension, and projection).

Much of the work for this project has been done in collaboration with the Information Quality Graduate Program at the University of Arkansas, Little Rock (UALR). Together with UALR, Qbase has investigated various methods for measuring the quality of sensor video streams. In addition, Qbase has researched and designed approaches for aggregating, storing, and processing this quality data and has investigated how to implement these approaches in the context of the related Persistent Surveillance Data Processing, Storage and Retrieval Project (*Task Order 005*, under the same contract).

2.0 INTRODUCTION

The focus of our research has been on understanding information quality from a variety of different perspectives in the context of evaluating persistent surveillance sensor data:

- Data quality vs. information quality
- Objective measures vs. subjective measures
- How objective and subjective quality measures affect the aggregate quality of the information
- Understanding and measuring the value of information
- Understanding how the quality and value of information is affected by the quality of the data from which the information is derived

We applied this research to develop approaches for analyzing, storing and communicating the quality and the value of information derived from persistent surveillance sensing activities. This included formats for describing sensor characteristics and sensor data streams and methods of enhancing sensor data with additional quality metadata. As part of this effort, we also developed a simulator using Matrix Laboratory (MATLAB) that allowed us to take previously recorded sensor data and degrade it in a variety of ways to evaluate its effect on the quality and value of the information being produced.

Throughout the course of the year we participated in a number of briefings that led us down the path of evaluating not only the quality of the data and information associated with sensor data streams but also the value of that information. One of the consistent themes we heard during these briefings was that image analysts, who are typically high school graduates with a minimal amount of experience, are tasked with reviewing large volumes of sensor data to determine whether or not there is anything of interest within the data. In addition, there is more sensor data being generated than there are analysts to review the data. As a result, there is an ever increasing chance that some items of interest might be overlooked or completely missed by the analyst. This situation leads to the need for more sophisticated tools that can assess, filter, and communicate only that data that is of most value to the analyst in achieving their mission.

In order to build tools that meet this need, it is first necessary to understand what factors are involved in determining the value of information and then providing a metric or metrics for assessing that value. Because the value of information differs based on the needs (or mission) of the analyst, it is important that those parameters be evaluated as well. This may mean that for each mission there is a different set of parameters and metrics that must be evaluated to determine the overall value of the data being analyzed. As part of this analysis, we evaluated information quality in the context of the process of situational awareness described by Mica Endsley in [24] as:

> …the **perception** of the elements in the environment within a volume of space and time, the **comprehension** of their meaning, the **projection** of their status into the near future, and the **prediction** of how various actions will affect the fulfillment of one's goals.

For the purposes of mapping our information quality research to this process we split the perception stage into two stages (Sense and Perceive) and we have not yet attempted to address the prediction phase. In order to define these stages more precisely we've drawn upon a number of different descriptions [24, 25, 26] to come up with the following:

- Sense – Capturing sensor measurements from the environment using one or more sensors
- Perception – Transforming sensor measurements into a set of facts (e.g. detecting events, identifying relationships) that describe the situation.
- Comprehension – Matching the set of known facts to previous situations to determine what activity is actually taking place.
- Projection – Envisioning the outcome or end-result of the situation based upon previous experience with similar activities.

The following table lays out, for each of these stages, some examples of the sensor processing tasks that would be performed as well as the data quality attributes that we would expect to collect:

Table 1: Situational Awareness Model

	Sense	Perception	Comprehension	Projection
Sensor Processing Tasks	• **Raw Sensor Data** • **Time/Location Metadata** • **Sensor Orientation Metadata**	• Georegistration • Target Detection • Event/Relation-ship Detection • Human Annotations	• Activity Detection • Analysis of Previously Detected Events and Outcomes • Human Annotations	• Projected Outcome • Analysis of Previously Detected Activities and Outcomes
Information Quality	• **Data Integrity** • **Time/Location Accuracy** • **Image Quality** • **Completeness**	• Registration Accuracy • Detection Certainty • False Alarm Rate • Trust Level • Convenience	• Detection Certainty • False Alarm Rate • Completeness • Relevance • Trust level • Usefulness	• Probability of Projected Outcome Occurring

During the "Sense" stage, we are capturing data from the environment using sensors. Sensors have varying levels of accuracy and integrity based on environmental conditions (time of day, weather, etc.), communication methods and protocols, compression and frame rates, etc. These are all data quality metrics that can be objectively measured either by the sensor, itself, or by the component of the system that is receiving the data.

During the “Perception” stage, the data is gathered and analyzed to determine the facts about the situation. For example, “What is the field of view of the sensor?” For airborne optical sensors attached to Unmanned Aerial Vehicles (UAVs) or piloted aircraft, this means understanding the location (latitude/longitude/altitude) and orientation (heading/roll/pitch/yaw) of the aircraft, and the orientation (pan/tilt) and optical characteristics of the sensor in order to determine its field of view. Each of these sensor measurements will have their own quality characteristics which must be taken into account when determining the overall accuracy of the field of view calculation.

Other facts that are relevant to this stage include the determination that a certain event is taking place or that a certain relationship exists. The algorithms that detect events or relationship have a certain probability of detection, probability of false alarm, and probability of missing a detection (Signal Detection Theory (SDT) [27] is commonly used to characterize these probabilities). These probabilities must be captured and propagated along with the data that describes the events/relationships that were detected in order to provide downstream algorithms and decision makers with the information required to understand the quality and value of the data.

We also have to determine whether or not to trust the data being provided by the sensor. Trust at this level could be based on the source of the data (i.e. is it coming from a trusted source?) or it could be based on other sensor integrity measures such as the reliability or past performance of the sensor. If the data provided by the sensor is not in a format we understand then it must be converted or discarded. Our ability to use this data and how difficult/reliable it is to translate the data into a format we can use is referred to as convenience. For instance, if we are sensing speech or text data that is in a different language than that of the analyst, the convenience factor will be lower than if it is in the native language of the analyst.

During the “Comprehension” stage, the facts are analyzed to determine whether any activity or activities are taking place that might be of interest. The algorithms that detect activities, similar to those that detect events/relationships will also result in probabilities of detection, false alarms and missed detection based on SDT that must be captured and propagated along with the data. In addition, the information produced in this stage of assessing the situation is evaluated for relevance and usefulness in the context of the mission. If the data is not relevant or useful, it is of low value and should be dropped from the data stream so as not to further utilize valuable computing or human resources.

During the “Projection” stage, the activities detected are evaluated to determine if a potential problem or threat exists. This could be based on the outcomes of similar activities that have occurred in the past or the analyst’s experience based on all of the facts, events, relationships and activities that comprise the current situation as well as the quality metrics generated in the prior stages. It is at this point that understanding the quality of the information being provided is essential in assisting the analyst or decision maker in projecting what the most likely outcome might be.

3.0 METHODS, ASSUMPTIONS, AND PROCEDURES

3.1 Requirements for the Layered Sensor Domain

3.1.1. Motivation

This section of the document addresses the issue of Information Quality in the layered sensor networks. Current surveillance systems use multiple sensors and media processing techniques in order to record/ detect information of interest in terms of events. Assessing Quality of Information is an important task as any misleading information may lead to suspicion and uncertainty to the decision makers. There is a need to evaluate the quality of streaming information in real-time, within the data stream as well as to perform the thorough forensic analysis such as different types of statistical analysis, historical trending of collected data, as well as probabilistic forecasting.

3.1.2. Challenge

The concept of Information Quality is not really well defined in persistent surveillance sensor networks. Bovik, et al. [1] provided several concepts for identifying Information Quality measurements in layered sensor networks. Very little was done for the video streaming data, which is of primary interest for the military surveillance systems. While there is a lot work done in terms of different quality algorithms for static images and even videos there is not enough information and publications addressing overall quality of information, quality control and monitoring of data stream information in surveillance systems, both real time and forensic.

3.1.3. Objectives

- Conduct and support research in the application of established information quality principles and methods, data integration, and data visualization in order to optimize the value of information obtained from layered sensor systems supporting persistent surveillance operations.

3.2 Research and Development Plan

- Conduct literature search for publications in scientific and technical research journals, conference proceedings, and other venue in order to document and build on existing knowledge of information quality concept in persistent surveillance sensor networks.
- Continue coordination with UALR the research and development of the objective and subjective metrics for video streams.
- Develop a conceptual model of organization of metadata within the data stream in accordance with Defense Advanced Research Project Agency (DARPA) situation awareness framework::Perception, Comprehension, Projection and Prediction-Quality of Information and Value of Information structures.
- Develop Quality of Information and VOI classes and attributes.
- Develop Quality of Information and VOI services within Persistent Sensor Storage Architecture (PSSA).
- Continue working on Aggregated Quality Score (AQS), Trust model and other information quality concepts.

- Continue integration with the PSSA and develop a demo with more or less realistic scenario. (Task 005).

3.3 Information Quality Processing, Propagation, and Storage

3.3.1. Problem Statement

The major problem of the data streams is a huge amount of data to be stored, processed and analyzed. Task Information Quality 006 was related to *Task 005* – Persistent Architecture. While Task 005 was intended to develop an architecture of ingesting, propagating and storing data obtained from layered sensors, the goal of *Task 006* was to develop a composable and flexible framework of enriching sensor metadata with quality indicators and to propagate and store these quality metrics along with the data stream without overloading the system in real-time and in forensic mode.

The Qbase team worked closely with UALR since UALR Information Quality department is well recognized in the Quality of Information area. UALR shared their quality algorithms and experimental setup with Qbase. Since UALR was working primarily with video streams and image processing principles, Qbase made a decision to create a composable framework of processing, propagating and storing quality metrics for the video data streams. In addition, Qbase has already acquired some AFRL data collect pieces such as Columbus Large Image Format (CLIF) data from the 2006 and 2007 Columbus data collects: Large Area Image Recorder (LAIR), Columbus Surrogate Unmanned Aerial Vehicle (CSUAV) and Ground Camera, and Video Verification of Identity (VIVID) data. In Phase II of the current project, a DARPA video data set was downloaded, incorporated into the Qbase demonstration software, and analyzed in accordance with developed quality framework and quality metadata metrics.

3.3.2. Standards for Machine Encoding of Sensor Data

Sensor Model Language (SensorML): A new way of storing, propagating and enriching data stream with metadata was proposed by introducing new Extensible Markup Languages (XML) specifically for sensor information. SensorML was developed by Dr. Mike Botts (University of Alabama, Huntsville) under the auspices of the International Committee for Earth Observing Satellites. The goal was to be able to exchange information between Location Services Clients (LSC) and location servers: http://www.ogcnetwork.net/SensorML Sensor Web Enablement (SWE) activity of the Open Geospatial Consortium (OGC) defines interfaces and protocols to access sensors over the Web. The following are the five foundational components:

- SensorML –The general models and XML encodings for sensors
- Observations & Measurements (O&M) – The general models and XML encodings for sensor observations and measurements
- Sensor Observation Service (SOS) – A service by which a client can obtain observations from one or more sensors/platforms
- Sensor Planning Service (SPS) – A service by which a client can determine collection feasibility for a desired set of collection requests
- Web Notification Service (WNS) – A service by which a client may conduct asynchronous dialogues with other services

SensorML has been designed for the following purposes, specifically to:

- Provide general sensor information in support of data discovery
- Support the processing and analysis of the sensor measurements
- Support the geographical location of the measured data
- Provide performance characteristics (accuracy, threshold, and so on)
- Archive fundamental properties and assumptions regarding sensor
- Support rigorous geographical location and mathematical models
- Apply to in-situ or remote sensors
- Support stationary or dynamic sensors

Information provided by the SensorML includes observation characteristics, physical properties measured (radiometry, temperature, concentration, and so on.), quality characteristics (such as accuracy, precision) which is especially valid for the current project, response characteristics (spectral curve, temporal response, and so on), geometry characteristics, geometric and temporal characteristics of sensor and sample collections (such as scans or arrays) that are required for metric exploitation and so on. It can also include description, documentation, and overall information about sensor, history, and reference information. All this means that SensorML has capabilities for enriching sensor data with quality metadata so that it can propagate along with the data stream and be stored along with the data stream easily. SensorML, for instance, was utilized for information storage and exchange by the Persistent Universal Layered Sensor Exploitation Network(PULSENet) (Northrop Grumman) [2]. Examples of SensorML are given on its website: http://www.botts-inc.net/vast.html

UncertML: The Uncertainty Markup Language (UncertML) is an XML schema for describing uncertain information and is capable of describing a range of uncertain quantities. Its descriptive capabilities range from summaries, such as simple statistics (e.g. the mean and variance of an observation), to more complex representations such as parametric distributions at each point of a regular grid, or even jointly over the entire grid. UncertML is XML encoding for the transport and storage of information about uncertain quantities, with emphasis on quantitative representations based on probability theory.

Typically most data contains uncertainty, arising from sources which include measurement error, observation operator error, processing/modeling errors, or corruption. Processing this uncertain data (typically through models, which can introduce their own errors), propagates uncertainty, and often unpredictably.

The ability to optimally utilize data requires a description of its uncertainty which is as complete and detailed as possible, and in the geospatial context, this characterization and quantification is particularly crucial when data is used for spatial decision making. Thus there is a well- recognized need for Geographic Information Science (GIS) frameworks which can handle and ‘understand’ incomplete knowledge in data inputs, in decision rules and in the geometries and attributes modeled.

A substantial literature exists on mechanisms for representing and encoding geospatial uncertainty and its propagation. However, no framework yet exists to describe and communicate uncertainty, either in Geographic Information (GI) data or more generally, in an interoperable manner. That is why UncertML was proposed to bridge this gap. For instance, as it is proposed in [3] UncertML can be utilized for a description of every pixel of a Geography Markup Language (GML) Rectified Grid. Every pixel can contain an UncertML Uncertainty as its value. This could be a Gaussian distribution, representing variance around the mean, or any other defined distribution.

Figure 1, below, shows an example of standard deviation encoded in UncertML [4]. Due to the soft-typed approach of UncertML all simple statistics will appear identical in structure. What separates a ‘mean’ from a ‘median’ is the Uniform Resource Identifier (URI), and definition upon resolving, of the definition property yielding a concise, yet flexible solution. Assuming the existence of a dictionary containing definitions of the most common statistics, only the URI is needed in order for an application to ‘understand’ how to process the data.

```
<un:Statistic definition="http://dictionary.uncertml.org/statistics/standard_deviation">
        <un:value>12.08</un:value>
</un:Statistic>
```

Figure 1: A Standard Deviation Encoded in UncertML

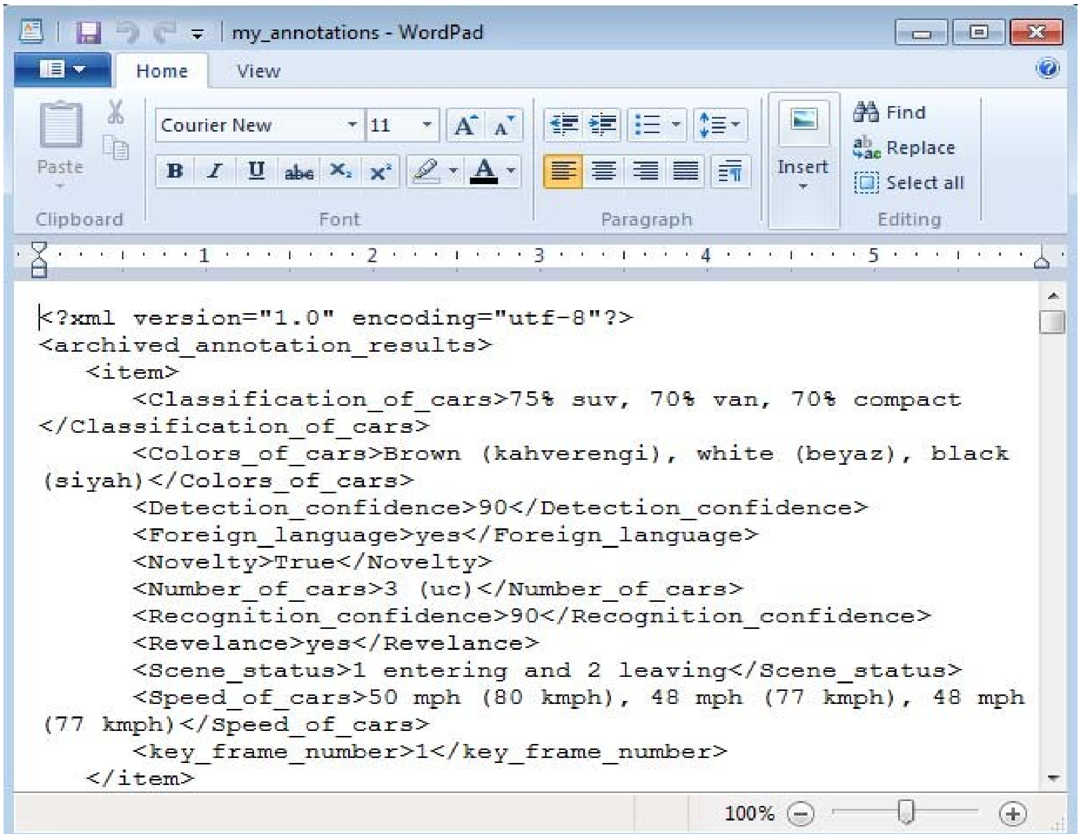
Summary: Based on this research performed in Phase 1 of the project, we developed a simpler XML format for adding and propagating the proposed quality metadata to the sensor data stream. Although we did not use SensorML/UncertML for our prototyping it is possible to transform this format into SensorML/UncertML XML, if required.

In Phase II we selected several data sources to demonstrate the proposed concept. One of the data sources –VIVID (see description below). VIVID metadata is originally recorded in the Transducer Markup Language (TML) format developed at AFRL. We developed a capability of reading VIVID data and adding quality metadata to it. An example is shown below in Figure 2.

It demonstrates how added annotations such as number of cars on the scene, car description, and new in Phase II subjective metrics such as Novelty, Relevance, and objective metrics such as detection and recognition confidence can be added as additional metadata metrics and then propagated with the data stream and retrieved for the analysis.

NOTE:

The explanation of these metrics is given in Section 3.7.



```
<?xml version="1.0" encoding="utf-8"?>
<archived_annotation_results>
   <item>
      <Classification_of_cars>75% suv, 70% van, 70% compact
</Classification_of_cars>
      <Colors_of_cars>Brown (kahverengi), white (beyaz), black
(siyah)</Colors_of_cars>
      <Detection_confidence>90</Detection_confidence>
      <Foreign_language>yes</Foreign_language>
      <Novelty>True</Novelty>
      <Number_of_cars>3 (uc)</Number_of_cars>
      <Recognition_confidence>90</Recognition_confidence>
      <Revelance>yes</Revelance>
      <Scene_status>1 entering and 2 leaving</Scene_status>
      <Speed_of_cars>50 mph (80 kmph), 48 mph (77 kmph), 48 mph
(77 kmph)</Speed_of_cars>
      <key_frame_number>1</key_frame_number>
   </item>
```

Figure 2: XML Format for Storing and Propagating Quality Metadata

3.4 Highlights of Phase I – Concept of Aggregated Quality Score

3.4.1. Concept of Trust Factor

The concept of Trust in sensor surveillance networks is given in several publications. One example is given in [5].

$$\begin{aligned} &Trust = Predictability + Dependability + Faith + Competence + \\ &\quad Responsibility + Reliability \end{aligned} \tag{1}$$

This equation demonstrates that Trust factor in sensor data streams is a combination of multiple factors. It definitely depends on the image or video quality of sensor data. In the above equation video quality can be represented by Dependability or Reliability variables. During Phase I of this research project, we investigated and evaluated the concept of aggregated video quality score based on the sum of weighted objective quality metrics. One of the examples is described in [6]. The regression model of independent variables-objective quality metrics can be run against dependent variable-subjective quality score. Based on the model, the objective quality metrics (noise, motion blur, blocking artifact, compression, resolution, etc) will have their weights calculated and those metrics with a stronger influence on perceived quality will be weighted higher as compared to other parameters.

$$AQS = \sum_{b=1}^{k} w_b \times q_b \tag{2}$$

Where

> AQS – *Aggregated Quality Score* , q_b – are separate quality factors and w_b is the weight of the b_{th} quality factor.

First task was to obtain objective quality measurements such as image quality metrics: noise, Structural SIMilarity Index (SSIM), blur, and so on. The required algorithms and degradation scenarios were provided by UALR. As for the subjective measurements we used those that were publicly available online and are described below.

3.4.2. Data Selection – Phase I

In Phase I we used publicly available Irvine Valley College (IVC) [7] and Laboratory for Image and Video Engineering (LIVE) [8], [18], [19] databases that were created by the University of Texas where videos and static images were subjectively evaluated by the Video Quality Experts Group (VQEG) on the scale from 1 – 5.

The IVC Image database consists of 10 reference images with 235 distorted images: *Joint Photographic Experts Group* (JPEG), JPEG2000, Locally Adaptive Resolution (LAR) coded and blurred. LIVE image database uses ten uncompressed high-quality videos with a wide variety of content as reference videos. A set of 150 distorted videos were created from these reference videos (15 distorted videos per reference) using four different distortion types - Moving Picture Expert Group (MPEG)-2 compression, H.264 compression, simulated transmission of H.264

compressed bitstreams through error-prone IP networks, and through error-prone wireless networks. Distortion strengths were adjusted manually to ensure that the different distorted videos were separated by perceptual levels of distortion.

Each video in the LIVE Video Quality Database was assessed by 38 human subjects in a single stimulus study with hidden reference removal, where the subjects scored the video quality on a continuous quality scale. The mean and variance of the Difference Mean Opinion Scores (DMOS) obtained from the subjective evaluations, along with the reference and distorted videos, are available as part of the database. In addition to videos from LIVE database AFRL data collects CLIF 2006/2007 data, VIVID. CSUAV were analyzed to obtain objective and subjective quality metrics.

3.4.3. Objective Quality Metrics

- Videos from LIVE database and AFRL video data streams available at Qbase were processed to obtain objective measurements of referenced and distorted videos. The examples of calculated objective quality metrics for these videos, such as noise, blur, SSIM and S-SSIM metrics [9] are displayed below.

Data processing was done by several methods:

- Qbase simulator and UALR tool were used to degrade Cliff 2006 data on frame by frame basis and receive quality metrics values per frame.
- Moscow State University (MSU) Video Quality Measurement Tool version 2.6 was used to obtain movie average objective quality metrics (frame by frame calculation is done as well) for the reference and compressed movies from LIVE and IVC databases.

A few examples of Image Processing by different applications are shown next in Figures 3, 4 and 5.

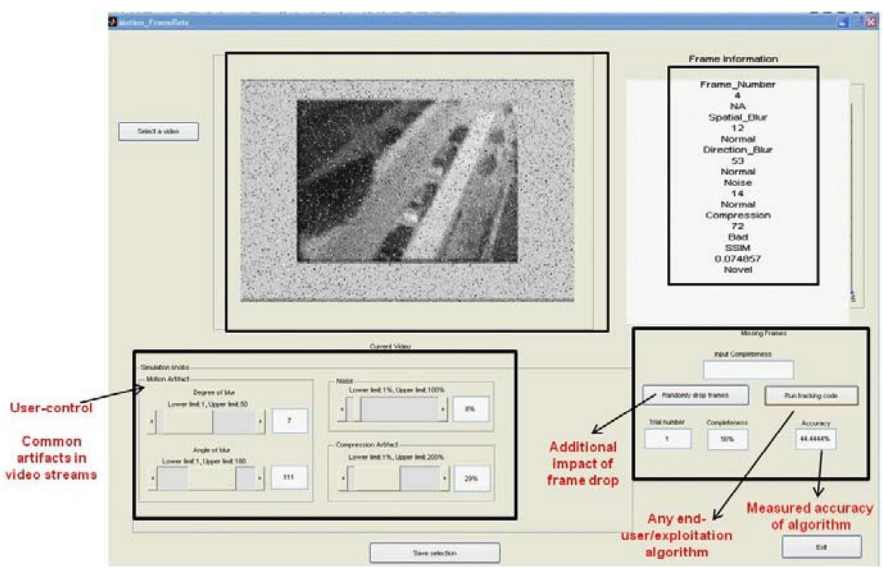


Figure 3: Qbase Simulator

Showing Controlled Amount of Degradation Added to CLIF 2006 Video

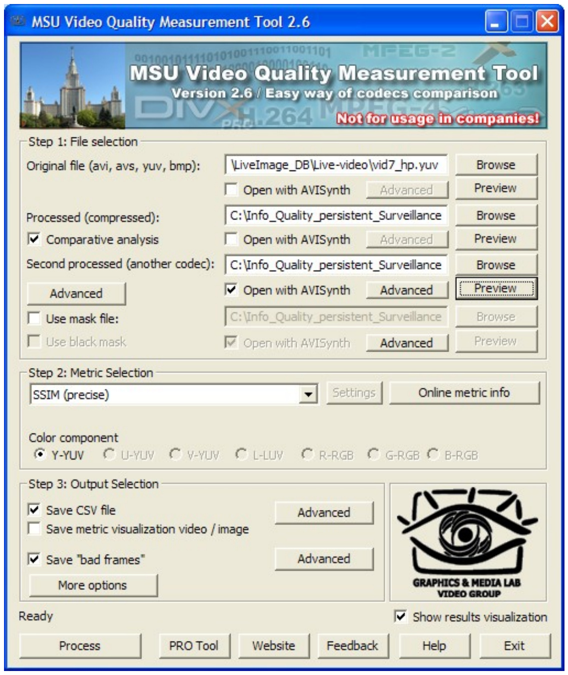


Figure 4: MSU Video Quality Measurement Tool

Video Data streams from Laboratory for Image and Video Engineering (LIVE) Video Database Used as Input Data

The MSU Tool provides the capability to calculate the average quality metrics such as noise, brightness, blur-beta, blocking artifact, SSIM, Peak Signal-to-Noise Ratio (PSNR), Mean Absolute Difference (MSAD), Delta, Frame Drop, Scene Change Detector for the entire movie, as well as frame by frame, see Figure 5 below. The metrics are described in the Appendix and on the MSU Video Quality Measurement Tool website [10].

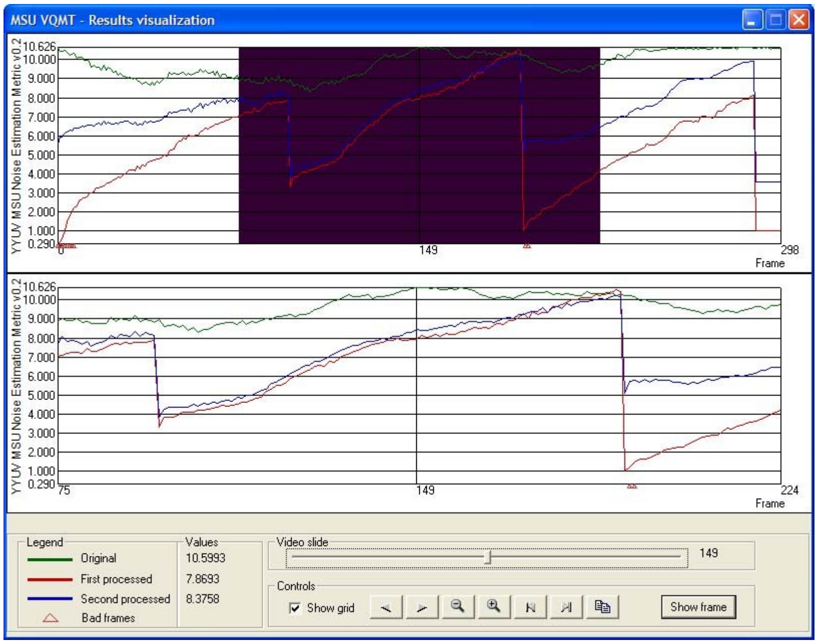


Figure 5: Examples of Metrics (Noise) Calculated for Three Videos from LIVE Database

Reference Video (Green), Blurred Image (Blue), Compressed Image (Red)

3.5 Subjective (Perceptual) Score

Subjective measurements were performed by using existing LIVE database Mean Opinion Score (MOS) provided by the Video Quality Experts group , MSU subjective measurement tool (demo) and Qbase simulator subjective interface. Examples of Subjective Score interfaces that were used to obtain subjective scores are shown below in Figure 6.

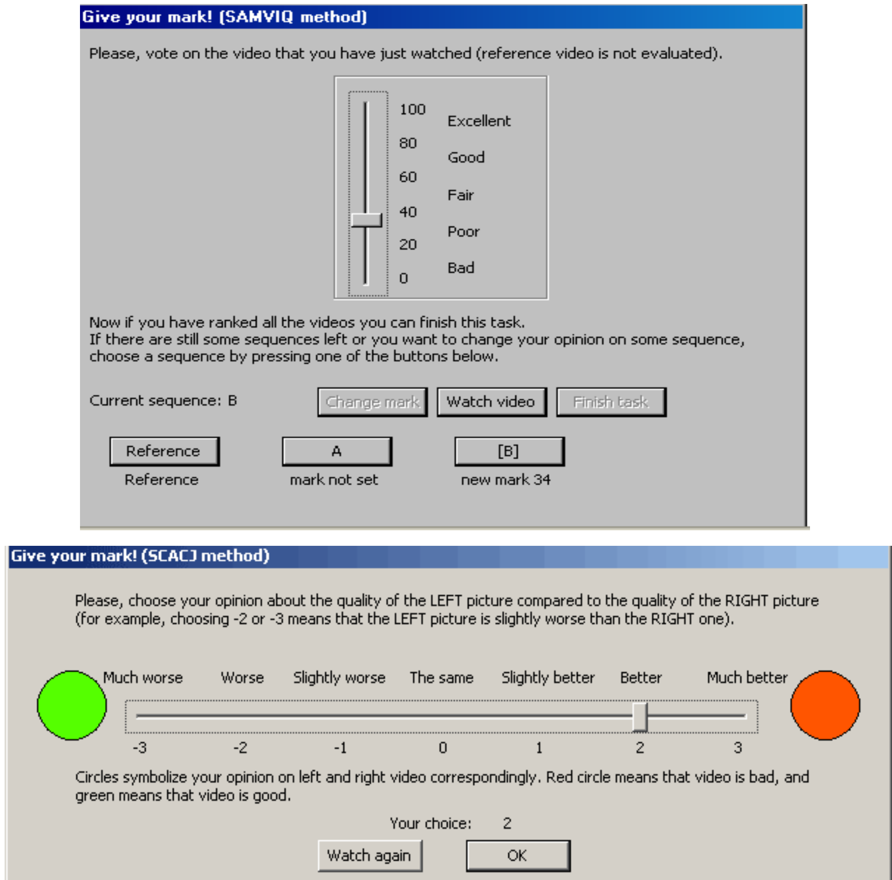


Figure 6: Examples of Different Subjective Measurement Tools

(MSU Perceptual Video Quality Tool)

Collecting suitable data for regression model could be a time consuming procedure. Only after data is been loaded into database, it has to be processed and analyzed, and only after this can a statistical analysis be performed and a corresponding regression model developed. Then, the Trust Factor/AQS model with weight factors as calculated can be reused for the new chunk of data stream, assuming that it was collected within the same surveillance system. It can be done in both forensic and real-time modes. This procedure allows monitoring the quality of the coming stream.

3.6 Regression Model

Collected data was used to build the Linear Regression Model (LRM) to calculate an AQS. Below is an example of calculated AQS for the analyzed video stream:

$$AQS = 3.01+0.02*avBlock-0.01*avNoise- +1.12*SSIM-0.09*avBlurr \quad (3)$$

It is evident that SSIM is the most decisive variable in the Video Quality Evaluation. However, the metrics for Trust factor may include various categorical values, for instance, such metric as data integrity – whether the data was manipulated or not, if there was any spam added, etc. Or such metric as data usefulness can have several levels- useful,-non-useful, and so on. In terms of Trustful or non-trustful information it makes sense to use a probabilistic approach- to evaluate a probability of whether you can trust or not trust information that you collected. Probabilistic model will be more useful in data fusion applications where the data from independent sources will be combined together to evaluate the Trust Factor of collected data stream. That is why we suggest to following model:

$$Probability\ (Trusted\ Event) = \frac{1}{1+e^{-(B_0+B_1X_1+B_2X_2+\cdots B_nX_n)}} \quad (4)$$

Where for instance:

X_1 = Image Quality – one or several combined image quality metrics such as noise, blur, SSIM or S-SSIM, blocking artifact, etc. It can also be an Aggregated Image Quality Score described above.

X_2 = Completeness- % of missing frames in video stream

X_3 = Timeliness-measure of information being available at desired time

X_4 = Integrity-available information has not been manipulated, etc…

B_0, B_1 ... B_n = regression coefficients

The histogram in Figure 7 gives a concept of understanding the probabilistic nature of the binary logistic regression model above. The symbol used for each case designates the group (trusted or not trusted) to which the case belongs. The cutoff value is 0.5. So the cases with certain combination of quality metrics where the calculated probability is higher than 0.5 can be considered as trusted information while those for which the probability is less than 0.5 cannot be considered as fully trusted information.

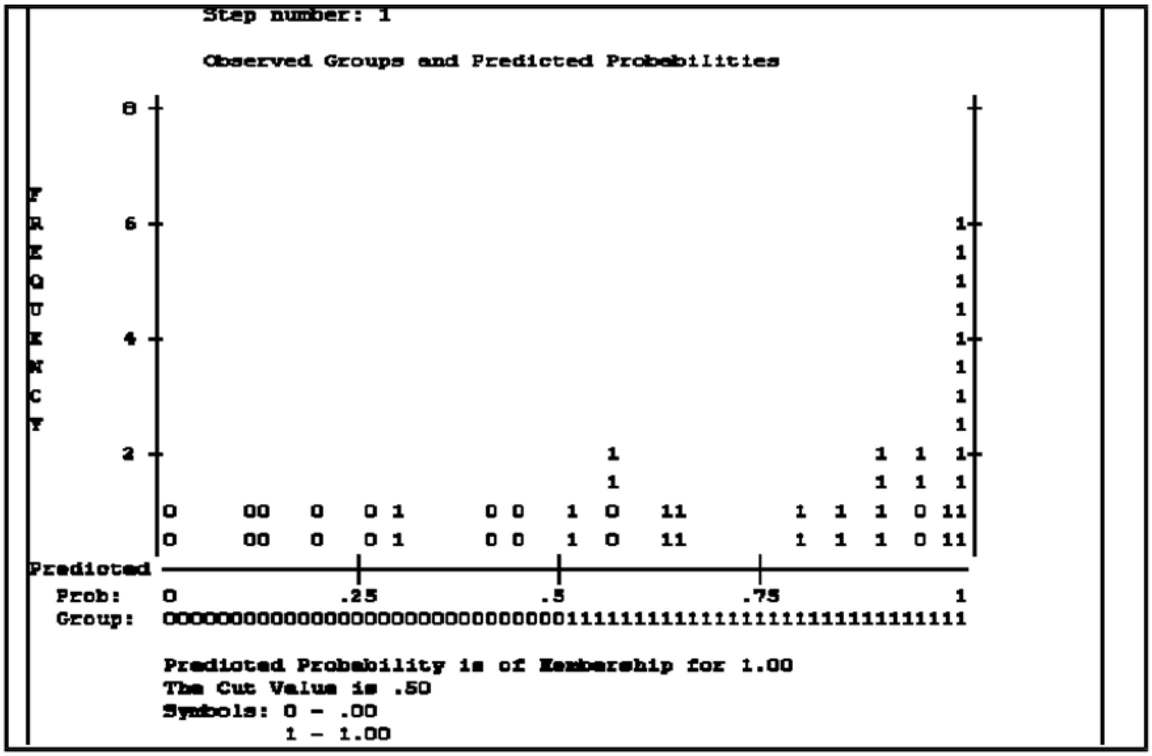


Figure 7: Histogram of Estimated Probabilities

A proposed regression model for Trust Factor has been developed in Phase II and is described in section 3.10.

3.7 DARPA Situation Assessment Principles – Perception, Comprehension, and Projection: Quality of Information and Value of Information

DARPA uses Endsley's situational assessment model described earlier to define the following major critical factors in the situation awareness or situation assessment:

- Perception – acquiring the available facts
- Comprehension – understanding the facts in relation to our own knowledge of such situations
- Projection – envisioning how the situation is likely to develop in the future, provided it is not acted upon by any outside force
- Prediction – evaluating how outside forces may act upon the situation to affect our projections.

Based on this concept we developed a new metadata structure that is in line with the principles of situation assessment. The quality of information will ultimately reflect upon its end-use. However, depending on the application, the end users may be interested in different pieces of information with different degree of quality. That is why, for Phase 2 of this research study, we proposed to split the quality metadata into 2 metadata models: one that relates to the inherent properties of information –

Quality of Information, and one that relates to the role of this piece of information in the context of its end-use-Value of Information. This approach was described in [11] and is presented in Figure 8.

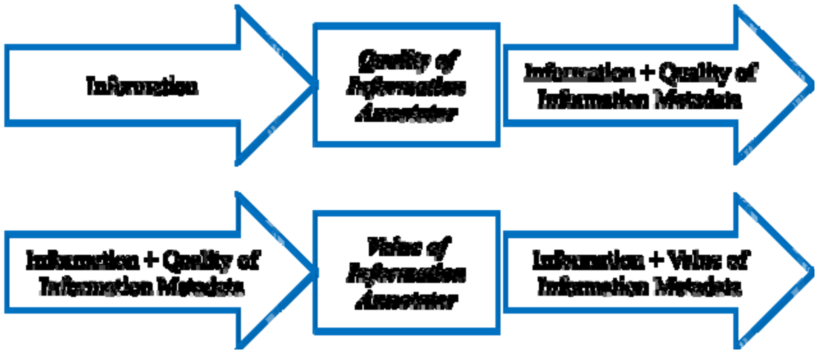


Figure 8: Proposed Organizational Structure of the Metadata

- Quality of Information Metadata - inherent characteristics of information that are independent of the specific application context in which the receiver will use the information and represents the ***Perception*** stage of situation assessment.
- Value of Information Metadata - *the utility of the information in an information stream when used in the application-specific* context of the receiver and represents the ***Comprehension and Projection*** stages of situation assessment

Quality of information (QoI) represents the body of evidence (described by information quality *attributes*) used to make judgments about the *fitness or utility* of the information contained in an information stream.

Value of information (VoI) represents the *utility* of the information in an information stream when used in a specific application context of the receiver.

Each metadata structure of Quality of Information and Value of Information can be represented by a collection of certain classes and attributes. These are proposed classes and related attributes for each of the models.

3.7.1. Quality of Information Data Structure

Proposed Quality of Information Metadata structure is presented in Figure 9.

Class	Attribute
Sensor Accuracy	SNR Temporal resolution Spatial resolution
Integrity	Owner Sensor Model Confidentiality Source Integrity
Timeliness	Latency
Format	Encoding type Compression ratio
Data Accuracy	AQS - Aggregated Video Quality Score Position Accuracy

Figure 9: Proposed Quality of Information Metadata Structure

The collection of QoI classes can include very different attributes depending on the sensor networks. We proposed several classes that can be more or less common to any sensor system available.

- The most common is Sensor Accuracy that can be described by Signal-To-Noise Ratio, Temporal and Spatial resolution attributes. These are typical sensor characteristics.
- Class Integrity can be described by the name of the Owner, Sensor Model, Source Integrity and Confidentiality if required for security purposes.
- Class Timeliness describes the ability to deliver information on time which can be defined by latency attribute.
- Class Format can be viewed as a representation of the quality of data, which measures quality related to the formatting of the information as data.
- Data Accuracy is the most important. It characterizes the data quality of incoming data stream.

One of the attributes of Data Accuracy – Overall Image or Video Quality, can be measured by the AQS developed in Phase I of this research project and described in earlier sections of this document (see Sections 3.4, 3.5, and 3.6 of this document).

3.7.2. Value of Information Metadata Model

Proposed Value of Information Metadata Structure is shown in Figure 10, below:

Class	Attribute
Trust	Data Accuracy Reliability of exploitation algorithms Sensor reputation Detection &Tracking confidence
Usefulness	Novelty (for object tracking) Relevance Time of The event/Latency Completeness
Convenience	Translation between measurement units for different members of the coalition
	Format Compatibility

Figure 10: Value of Information Metadata Structure

- The ***Trust*** class comprises the reputation of the source of the information, its objective quality, and the reliability of the source all as perceived by the receiver.
- The ***Usefulness*** class captures the usefulness of information in a specific context as determined by the receiver. The usefulness is assessed along four attributes, one indicating the level of novelty of the information received, a second measuring whether the information achieved is relevant for the needs of the receiver, the third expressing how timely the information is for the purpose of the receiver, and a fourth can be expressing the level of completeness of the information. The completeness of the information measures the degree by which the information at hand covers all that is needed by the receiver.
- The ***Convenience*** class captures how easy or difficult it is for the receiver to use the information and is assessed along three attributes. The format of the information, whether it is readily usable by the systems of the receiver or requires manipulation is assessed by the *Format Compatibility* attribute.

3.8 Data Selection – Phase II

In Phase II we worked with several datasets, VIVID and Video Image Retrieval and Analysis Tool (VIRAT) data. The VIVID data was collected at Fort Pickett and Fort Lee 2004. It is stored in TML format – which can be compatible with ***SensorML***, ***UncertML*** (see above). It consists of video frames and sensor platform metadata. The resolution of each clip is 640x480, rate is 30 frames per second. Filename format is given in a form V4VxZZZZZ_YYY.avi where Vx represents the sensor:

- V1: EO Daylight TV (DLTV)
- V2: EO DLTV Spotter

- V3: IR
- ZZZZZ 5 digit number representing the scenario
- YYY - 3 digit number representing a clip from a scenario

Figure 11 below is an example of VIVID dataset run through the simulator with added Information Quality attributes.

Figure 11: Original VIVID Movie with the Information Quality Attributes Described Above

The second dataset, VIRAT, includes high quality videos recorded from total 6 scenes, captured by stationary High Definition (HD) cameras (1080p or 720p). There may be very slight jitter in videos due to wind. Videos are encoded in H.264.

Each video clip will contain 1~20 instances of events from 6 categories: (1) person loading an object to a vehicle, (2) person unloading an object from a vehicle, (3) person opening a vehicle trunk, (4) person closing a vehicle trunk, (5) person getting into a vehicle, and (6) person getting out of a vehicle.

This dataset was selected also because it provides such data as scoring confidence that we substitute for detection and recognition confidence and used in Trust factor modeling. Detection scoring confidence is provided with the different samples of VIRAT data annotation files.

An example of one of the VIRAT scenes is presented in Figure 12 below:

Figure 12: One of the Scenes from the VIRAT Dataset

Here three instances are captured- green bounding box is drawn around the object #1- the person, purple bounding box is drawn around the second object- the car, and the red bounding box is drawn around the event (6) – person getting out of a vehicle.

3.9 Tracking and Detection

3.9.1. Video Segmentation Tool

Qbase has worked with UALR and incorporated the Video Segmentation tool developed by UALR into our demo software. This tool allows the user to manually enter the comments and observations based on frame by frame analysis. The results are demonstrated below in Figure 13.

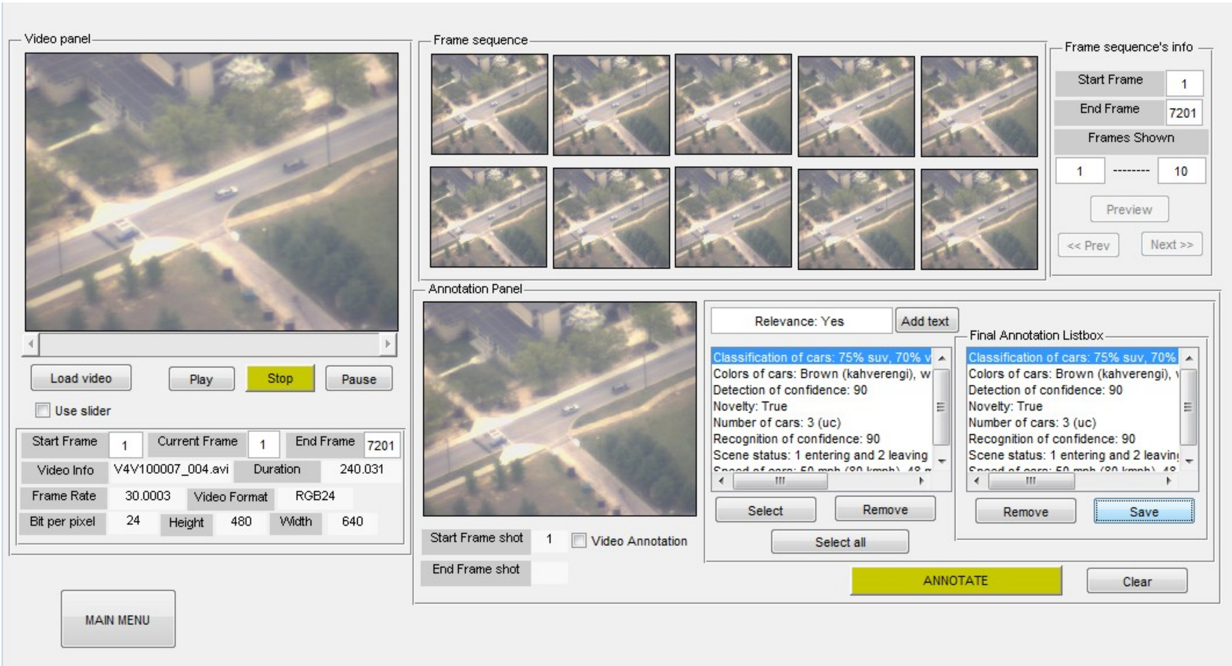


Figure 13: Video Segmentation Tool Developed by UALR

The annotation results were incorporated into XML metadata file that is propagated and stored along with the data stream. The resulting data table with such Usefulness attributes as Novelty and Convenience is presented below in Figure 14. The proposed metrics can eliminate tedious data and image analysis allowing to select only those frames where, for instance, “Novelty=True” and “Relevance=Yes.” This will significantly reduce the data load for image analysts.

The annotations were performed manually but they mimic the results of automated detection and recognition algorithms. These output results (quality metrics) that can be added to metadata structure and can be propagated and stored along with the entire data stream. They can be used in modeling the Trust or Usefulness factors of VoI Metadata Structure.

	A	B	C	D	E	F	G	H	I	J	K	L
1				Description								
2	ID	Time stamp	Frame Number	color	number	scene status	Speed	Classification	Detection Confidence	Recognition Confidence	Novelty	Relevance
3		1082473571700	1	Brown, white, and black	3	1 entering 2 leaving	50, 48, 48 mph	70% suv, 70% van, 70%compact	90	90	TRUE	yes
4		1082473573890	74	Black, brown, white and white	4	2 entering 2 leaving	40, 45, 45, 45	75% compact, 70% suv, 60% van, 75% compact	80	80	TRUE	yes
5		1082473589970	610	Black, white and brown	3	1 entering 2 leaving	40, 40, 40	85% compact, 80% suv, 60% van	75	80	FALSE	no
6		1082473592370	690	Brown, white, blue and red	4	1 entering 3 leaving	30, 20, 10, 10	70% suv, 60% van, 65% compact, 70% compact	70	70	TRUE	yes
7		1082473572607	937	Brown, white and blue	3	2 entering 1 leaving	50, 50, 50	70% suv, 70% compact, 60% van	70	70	TRUE	yes
8		1082473572765	1095	Yellow, brown, white and blue	4	2 entering 2 leaving	50, 50, 50, 50	80% compact, 60% van, 70% suv, 70% compact	80	80	FALSE	yes
9		1082473572958	1288	Brown and silver	2	2 leaving	50, 50	60% van, 80% suv	80	80	TRUE	yes
10		1082473573065	1395	Silver and blue	2	1 entering 1 leaving	40, 40	70% suv and 70% compact	70	70	TRUE	no
11		1082473573116	1446	Brown and blue	2	1 entering 1 leaving	40, 40	60% van, 80% compact	80	80	FALSE	no
12		1082473573182	1512	Silver and white	2	1 entering 1 leaving	40, 40	80% suv, 80% compact	80	80	TRUE	yes
13		1082473575552	1591	Brown, silver and blue	3	1 entering 2 leaving	40, 40, 40	60% van, 70% suv, 70% compact	80	80	TRUE	no
14		1082473576992	1639	Brown and silver	2	2 leaving	40, 40	60% van, 70% suv	80	80	TRUE	yes
15		1082473582992	1842	Brown, white and blue	3	1 entering and 2 leaving	40, 40, 40	80% suv, 80% van, 80% compact	80	80	TRUE	yes
16		1082473590882	2105	Brown, blue, white and white	4	2 entering and 2 leaving	10, 10, 15, 15	50% van, 60% compact, 60% suv, 70% compact	70	70	FALSE	yes
17		1082473594212	2216	Brown, white, blue and white	4	3 entering and 2 leaving	10, 10, 5, 10	40% van, 60% suv, 70% compact, 70% compact, 70% compact	70	70	TRUE	yes
18		1082473596492	2292	Brown, white, white, blue, silver, blue	6	4 entering and 2 leaving	0, 10, 10, 15, 15, 1	% van, 60% suv, 60% compact, 60% suv, 60% compact, 60% comp	60	60	TRUE	yes
19		1082473598382	2355	Silver, white, blue, white and black	5	3 entering and 2 leaving	10, 10, 10, 15, 15	50% van, 50% compact, 60% compact, 70% compact, 70% compact	60	60	TRUE	yes
20		1082473602522	2493	own, white, blue, white, black and silv	6	4 entering and 2 leaving	0, 20, 10, 10, 10, 1	% van, 80% suv, 55% compact, 50% compact, 50% compact, 60% s	60	60	FALSE	no
21		1082473603812	2536	hite, blue, white, black, silver and bro	6	5 entering and 1 leaving	0, 10, 10, 10, 10, 1	suv, 60% compact, 50% compact, 50% compact, 60% suv and 60%	60	60	TRUE	yes
22		1082473606902	2639	llow, blue, white, black, silver and bla	6	5 entering and 1 leaving	0, 20, 10, 10, 10, 1	% compact, 60% compact, 70% compact, 70% suv, 70% suv, 60% s	55	55	TRUE	yes
23		1082473610292	2752	Yellow, blue, silver and brown	4	4 leaving	20, 10, 10, 10	60% suv, 60% suv, 60% compact, 60% suv	55	55	TRUE	yes
24		1082473618602	3029	Blue	1	1 entering	40	51% suv	60	60	FALSE	yes
25		1082473620642	3097	Blue and brown	2	2 entering	40, 40	51% suv and 60% suv	55	55	TRUE	yes
26		1082473620658	3113	Silver, black and brown	3	2 entering and 1 leaving	40, 40, 40	%60 suv, 51% compact, 60% suv	55	55	TRUE	yes
27		1082473627168	3330	Black	1	1 leaving	40	51% compact	51	51	TRUE	yes
28		1082473633558	3543	Red and black	2	2 entering	40,40	80% truck and 60 % compact	51	54	FALSE	no
29		1082473635748	3616	Red and black	1	1 entering	40	80% truck	52	52	TRUE	no
30		1082473651318	4135	White and silver	2	2 entering	40,40	60% compact and 60% compact	51	51	TRUE	yes
31		1082473655338	4269	Silver	1	1 entering	30	60% compact	53	53	FALSE	yes
32		1082473656688	4314	Silver and brown	2	1 leaving and 1 entering	30, 30	80% compact and 60% van	53	53	TRUE	yes

Figure 14: Annotation Results

3.9.2. Tracking and Detection Information

Additional tracking and detection information from VIRAT dataset has been added to provide detection and tracking information to entire information data stream. The VIRAT dataset already has its own quality metrics such as scoring confidence, Precision, Probability of Detection, False Alarm Rate, etc., where:

- Precision is the ratio TP/D where D is the total number of detections (correct and incorrect) and TP is the number of correct detections.
- Probability of Detection is the ratio TP/T for every category, where T is the number of ground-truth activities in archive, and TP is the number of correctly detected activities matched to a member of T according to the activity-matching criterion.
- False Alarm Rate is the ratio False Positive/Normalizing factor (FP/NORM), where FP is the number of false positives whose detected activities do not match a member of T, and NORM is a normalizing factor based on the number of frames so that FP/NORM is in units of *activities per minute*

This data has been provided in VIRAT summary files along with the scoring confidence. This confidence was calculated with scoring software developed as a part of Computer Vision and Pattern Recognition (CVPR) Activity recognition Competition. The scoring confidence was calculated by comparison of the "participant's detection" versus ground truth specified by VIRAT data. This was done for objects (car, person, etc.) and for events as well [12] ,[13] and [14] .

We used scoring confidence data from the summary files to provide Detection confidence values that we interpret as objects detection confidence and Recognition confidence values that we interpret as event's detection confidence for Trust model calculations (both values were set equal to scoring confidence). This data is supposed to mimic the results of the automated detection and recognition exploitation algorithms that can be incorporated into PSS architecture in the future. Precision and Probability of Detection can be included in Trust calculation in the as well. The results are discussed in the section below.

3.10 Revised Trust Model

Beginning in Section 3.4 of this document, we suggested a concept of a Trust factor based on the AQS that was developed during Phase I of this research project and other quality metrics. In Phase II we've revised this model to include the trust class attributes associated with the VoI as described in Section 3.7.2 of this document. The probabilistic model was chosen to calculate Trust value for VIRAT video dataset.

Trust class attributes as it is demonstrated in the VoI section (objective quality metrics presented by AQS, detection and recognition confidence, Reliability of Algorithms) will have their weights calculated and those metrics with a stronger influence on perceived quality will be weighted higher as compared to other parameters. The binary logistic regression was chosen that predicts Probability of Trusted Event occurring. Subjective Trust variable was calculated on the % scale, from 0- 100%

$$Probability\ (Trusted\ Event) = \frac{1}{1+e^{-(B_0+B_1X_1+\ B_2X_2+\cdots B_nX_n)}} \qquad (5)$$

Where:

X_1 = AQS- a combination of image/video quality metrics such as noise, blur, SSIM or S-SSIM, resolution, etc.

X_2 = Completeness- % of missing frames in video stream

X_3= Reliability of Exploitation Algorithms

X_4 = Detection confidence

X_5 = Recognition confidence

B_0, B_1 ... B_n = regression coefficient

$$Log\left(\frac{Prob\ (Trusted\ Event)}{Prob(Untrusted\ Event)}\right) = -12.1 + 2.9 * AQS + 15.3 * DetectionConfidence + 11.1 * RecognitionConfidence - 2.0 * AlgortihmReliabilty + 3.1 * Completeness \quad (6)$$

In binary logistic regression modeling, the cases with certain combination of quality metrics where the calculated probability is higher than 0.5 can be considered as trusted information while those for which the probability is less than 0.5 cannot be considered as fully trusted information.

The dependent variable Trust was not readily available for modeling like the Mean Opinion Score from Video Experts Group that was used to model the Aggregated Video Quality Score. That is why it was evaluated by Qbase based on the goodness of objects' and events' detection and recognition. No wonder why detection and recognition confidence variables have the strongest weight in the above equation. Other independent variables such as AQS components and Completeness were modeled by degrading video quality using Qbase simulator. Algorithm Reliability variable was varied randomly between 1 and 0.75 just to be shown that this variable can be important in Trust factor analysis.

Similar approach can be used to calculate another VoI factor - Usefulness. Again, attributes such as Novelty, Relevance, Completeness and Timeliness can be composed into a regression model that will provide the dependent variable such as Usefulness.

This parameter may accept both quantitative and enumerated values, novelty, for example, may be an enumerated attribute with values: redundant, corroborative, incremental, new, or surprising!

The above model is just one proposed example of how to combine different objective quality metrics into the Trust factor. To continue with this model more data need to be analyzed and tested, with various scenarios and various subjective scores should be collected.

3.11 Phase II Video Analysis

In Phase II, additional Persistent Sensor Storage (PSS) services were developed to demonstrate the concepts of Quality of Information and VoI. They are listed below.

- Sensor Metadata Service – this service was added to existing PSSA to generate such static Quality of Information attributes as Sensor Accuracy, Integrity and Format. It also calculates Completeness metric that will be further passed to the Quality of Information Service.
- Quality of Information Service- this service gets image data from video image ingestor and calculates different metrics such as noise, blur, SSIM , resolution, etc.,

that are combined into AQS and sends them to VoI Service or directly to the dashboard.

- Value of Information Service- this service receives AQS from Quality of Information Service, Tracking and Detection confidence values (in this phase they are available from existing VIRAT data or added via UALR segmentation tool), Completeness, Novelty, Timeliness, Reliability, etc. to calculate Trust, Usefulness and Convenience correspondingly and sends them to the dashboard.

The schematic of these Services is displayed below in Figure 15.

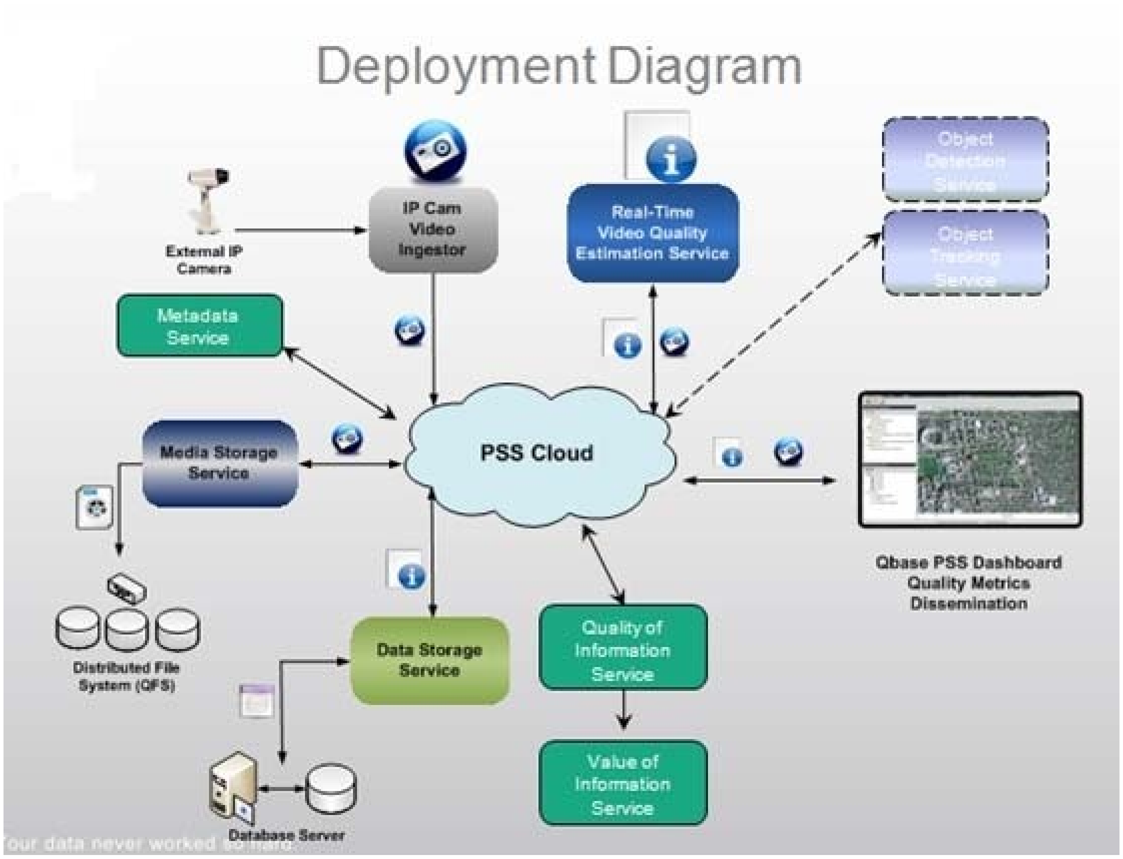


Figure 15: PSS Services to Support Info Quality

Metrics used to model Trust factor were obtained by analyzing VIRAT data. Again, similar to Phase I this video was analyzed as-is and with certain degradation added to it. The original and degraded videos are shown below. The controlled amount of noise and blur were added randomly. Also, random frames were dropped to add a Completeness metric, which was calculated as the number of frames that have been sent out.

Completeness= (Total # of frames- total # of dropped frames)/total # of frames

The resulting metadata is shown on the right side of the video, Figures 16 and 17 below.

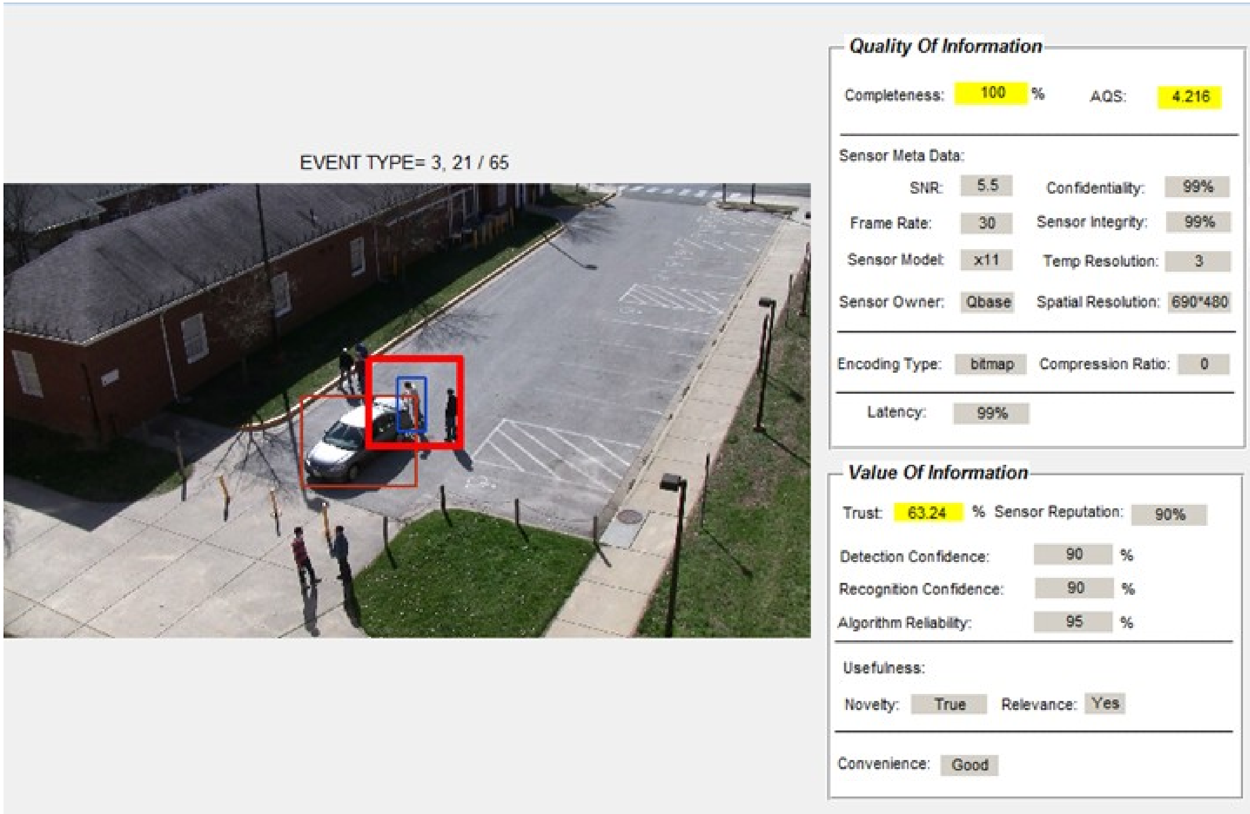


Figure 16: Original VIRAT Video

Figure 17: VIRAT Video with Degradation Added

3.12 Different Scenarios Analyzed in Phase II

Different scenarios were created in Qbase's simulator to demonstrate the capabilities of the developed architecture and the quality metadata that can be analyzed, propagated and stored along with the information data stream.

The first two scenarios were designed to demonstrate the effect of different objective image quality metrics on the overall AQS and correponding Trust Factor. Resolution and noise quality metrics were chosen to demonstrate how the combination of these metrics with different coefficients may influence the Image Quality and Trust factor.

Figure 18, below, demonstrate the combination of average noise and high resolution scenario. As you can see, the AQS is decent and equal 3.3 and Trust Factor is good enough, equal to 75%.

The objects and event surrounded by red bounding box are still trackable and can be detected and recognized.

Figure 18: Average Noise and High Resolution Scenario

The second scenario demonstrates the effect of low resolution - the video was degraded by adding a little noise and by change of resolution, Figure 18. As it is demonstrated, the situation is worse, although a very little noise has been added. The AQS and corresponding Trust Factor are lower, since it is impossible to detect many different objects such as people and event.

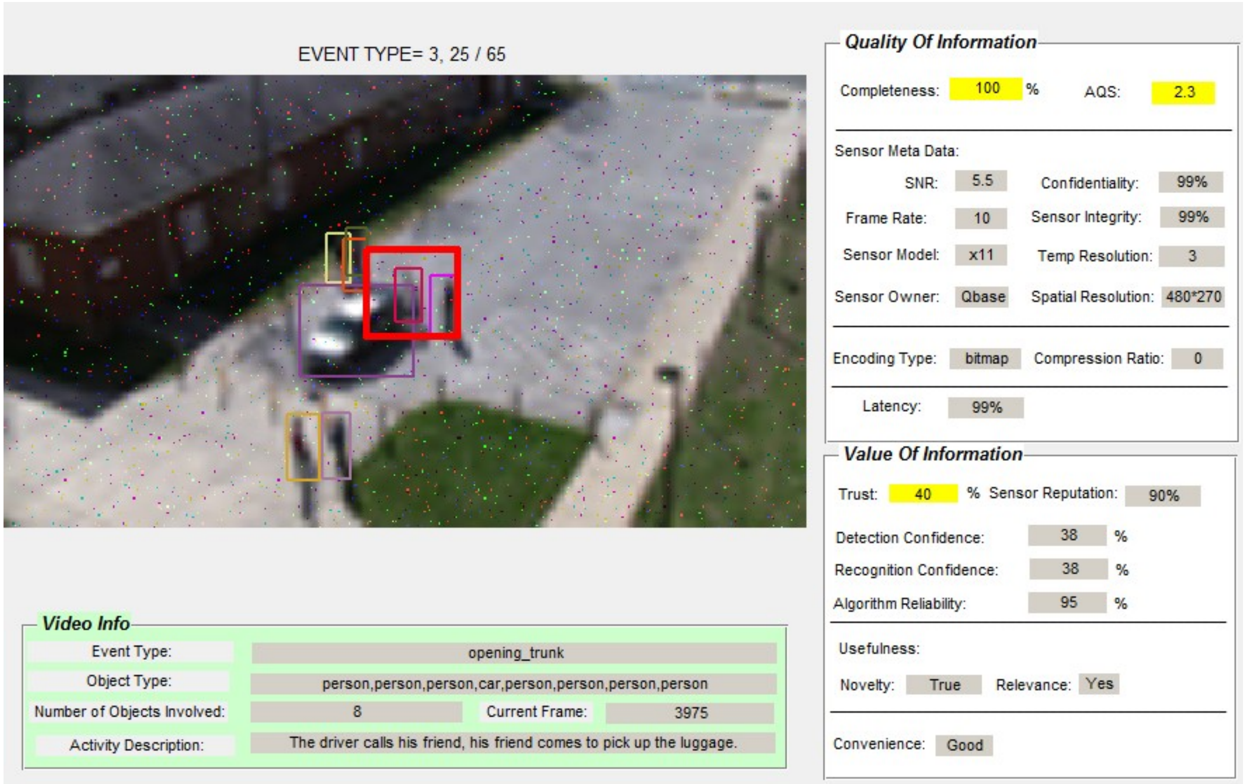


Figure 19: Low Noise and Low Resolution Scenario

The third scenario was created to demonstrate the effect of such quality metric as completeness (% of dropped frames). As an example, we tried to demonstrate how dropping frames will result in the uncertainty of the activity procedure.

We identified the "Suspicious" activity chain - as follows:

- Event #6 – Person #1 gets out of the car
- Person #2 comes to the car where Person #1 is
- Person #2 opens the trunk of the car (event type #3)
- Person #2 gets an "suspicious" object from a vehicle – event #2
- Person # 2 leaves the scene with an object from the vehicle

This "Suspicious" activity is demonstrated in Figures 20 through 24 (no degradation metrics have been added to the video).

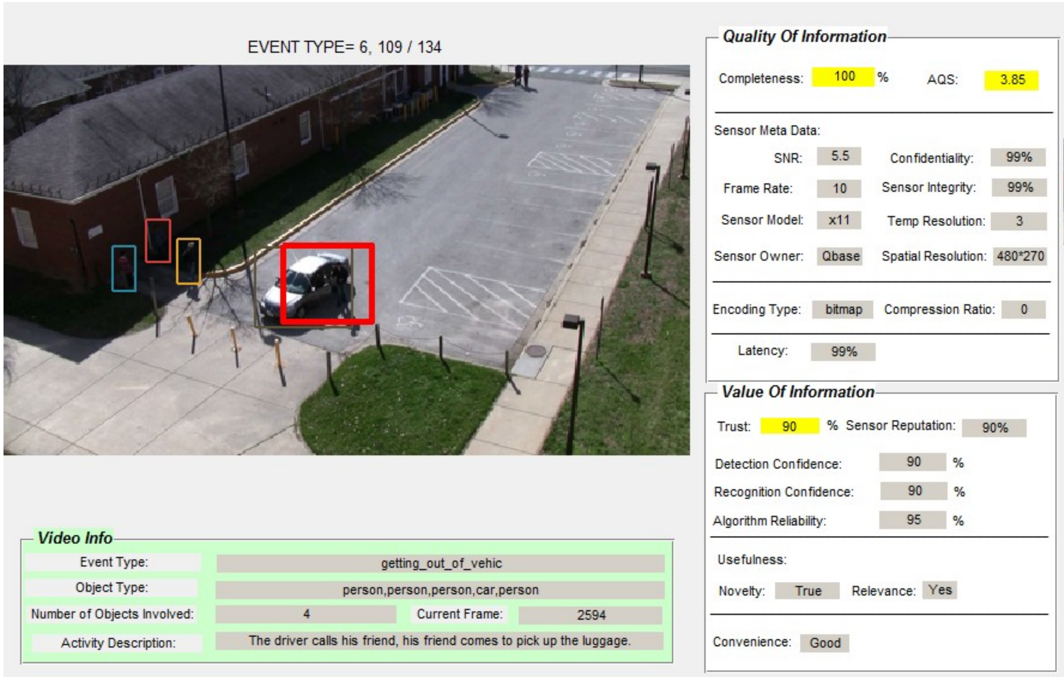


Figure 20: "Suspicious" Activity Non-Interrupted – Person #1 Gets Out of the Car

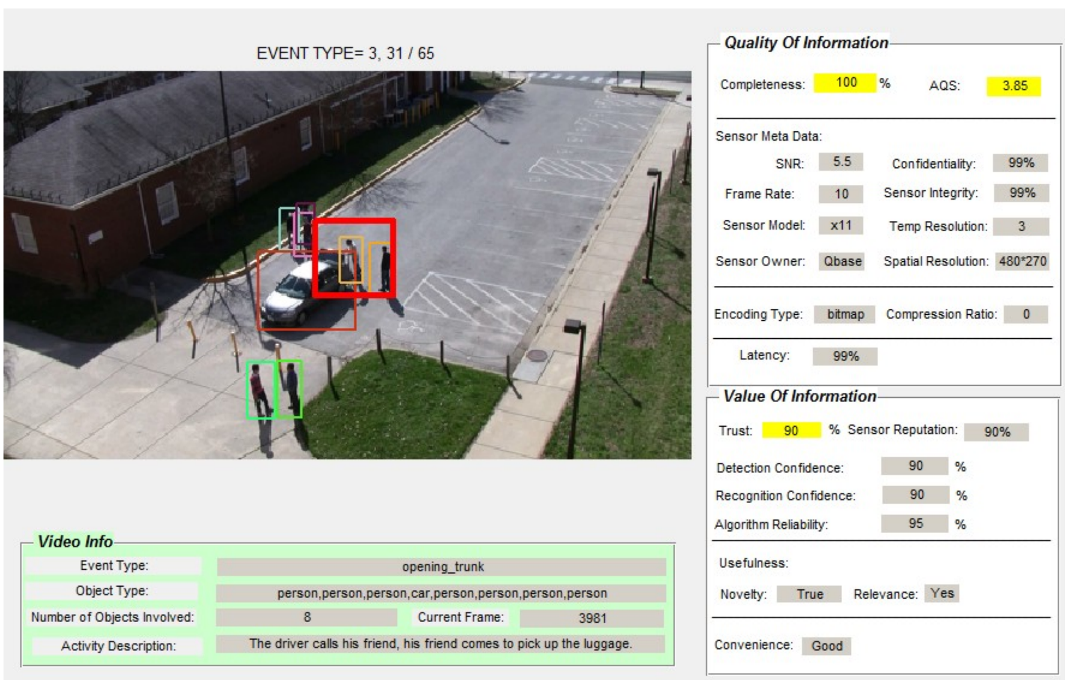


Figure 21: "Suspicious" Activity Non-Interrupted – Person #2 Comes to Person's 1 Car, Opens Trunk and Unloads the "Suspicious" Object

Figure 22: "Suspicious Activity" Non-Interrupted – Person #2 Carries a "Suspicious" Object Away and Person #1 Gets into the Car.

Now the frames between 2594 and 4378 are dropped. The result is shown in Figure 23.

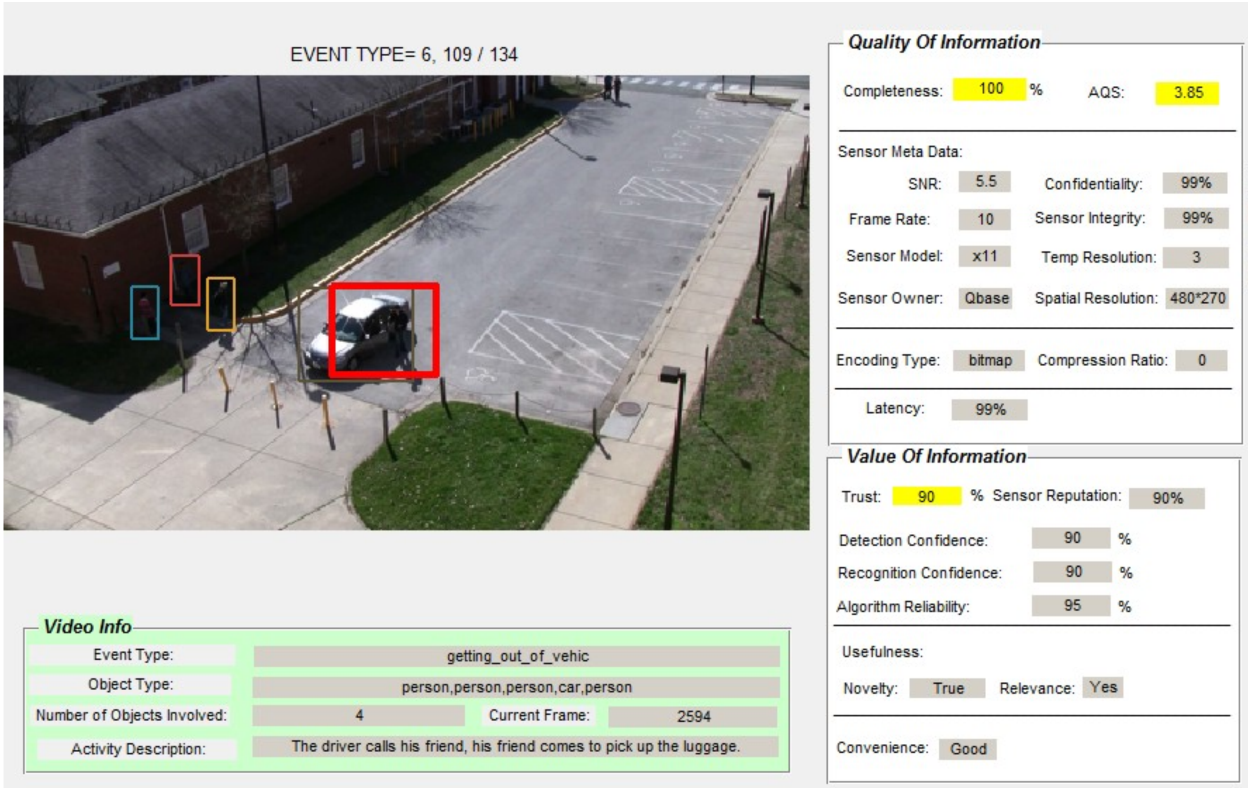


Figure 23: The Start of "Suspicious" Activity – Person #1 Gets Out of the Car, No Changes Here

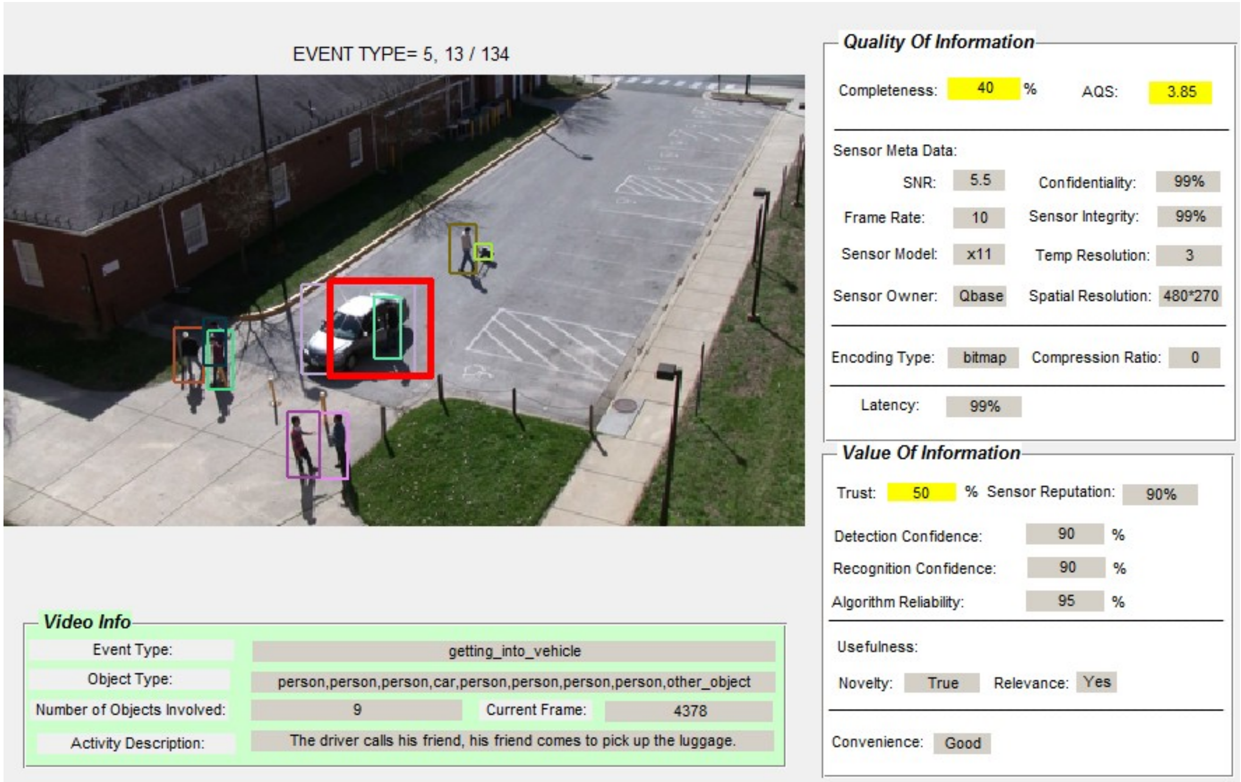


Figure 24: "Suspicious" Activity Interrupted – Person #2 is Carrying "Suspicious Object" Away

Figure 23 shows the previous frames where Person #2 meets with Person #1, opens trunk, and unloads the object from the trunk have been dropped. This time frame, 4378, comes right after frame 2594 (Figure 24). We are not sure now where Person #2 got that object if there is any connection to Person #1 and his car. The Completeness factor has dropped to 40%, and as a result of this, the Trust Factor has dropped as well.

3.13 Data Quality Processing within a Data Stream Management System (DSMS)

In persistent surveillance networks all kinds of information quality issues can occur: malfunctioning sensors, wrong sensors setups, wrong sensor calibration, incomplete data - missed video frames for instance confidence level of data is not acceptable, data is not accurate, data has been delayed, and so on. That is why it's important to have a quality control system in place in real time and quality metrics - the additional metadata should be able to propagate and stored along with the data stream itself. The proposed "Jumping Window" architecture below allows enriching of the sensor data stream with quality metadata without overloading the entire system. The concept of the Jumping Window architecture was proposed originally in [15] for residual lifetime of a truck's engine; we decided to apply a similar concept to video data streams.

Jumping Window/Sampling architecture was proposed and developed is Phase 1 as an extension of the conventional DSMS and is presented in Figure 33. (The idea is to propagate the data measurement with quality information for each Data Quality (DQ) dimension (shown in gray) with the same stream rate as the measurement stream (shown in white).

Each measurement attribute stream is divided into an unlimited number of windows with a given size *s* containing sensor data itself and quality metadata (examples are -Completeness, Aggregated Quality Score and Trust), Figure 25. Each window is identified by its starting point $Frame_{begin}$ and consists of *s* measurement values of a certain attribute. The window contains one value for each metadata attribute .The number of data quality attributes can vary. The window size *s* can be defined independently for each stream attribute.

As a result the quality metadata is not sent together with every single data item, but rather window-wise for each DQ dimension. The data volume is reduced significantly by aggregating the quality metadata for each attribute within window of the given size s_i starting at $Frame_{begin}$. This prevents the real time data stream and the storage from being overloaded. Aggregation functions can be flexibly determined for each DQ dimension depending on the application.

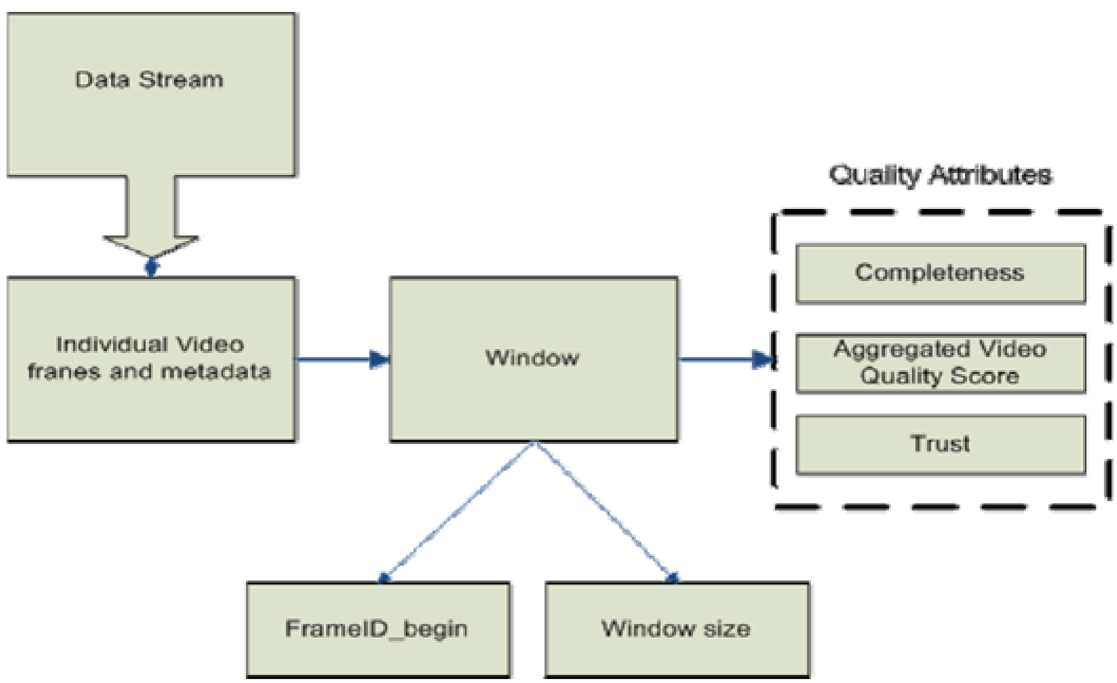


Figure 25: Jumping Window Architecture for the Propagation of Data Quality (DQ) [11]

The Jumping Window concept was implemented using the Qbase Simulator. Quality metrics were processed and aggregated for every ten video frames. An example is provided below in Figure 26, next.

Quality_ID	Quality Metric name	Frame_begin	Frame_end	Aggregated Quality Value	Aggregation Algorithm
1	Noise	1	10	43.33	Linear Average
2	Blur	1	10	6.5	Linear Average
3	SSIM	1	10	75.4%	Linear Average
4	Completeness	1	10	98.3%	Linear Average
5	AQS	1	10	4.34	Weighted
6	Detection confidence	1	10	87%	Linear Average
7	Tracking confidence	1	10	90%	Linear Average
8	Timeliness	1	10	100%	Linear Averaged
9	Trust	1	10	7.9	Weighted
10	Novelty	1	10	True	Random
11	Relevance	1	10	Yes	Random

Figure 26: Quality Metrics Aggregated Every Ten Frames

3.14 Metamodel Extension in Database Management System (DBMS)

Data quality can be considered as a new dimension in the relational metamodel. Every column in a relational table is enhanced with d data quality characteristics (DQ dimensions). Following the concept of not overloading the storage system, the extension metamodel is developed so that data quality information is not stored for every measurement value v_{ij} The Jumping Window aggregated metrics are stored in the database in a separate table that is mapped to the data stream.

The concept of MetaMapping a Jumping Window in the database is shown below in Figure 27. The measurements of the data stream refer to the respective columns in the DQ table. For each incoming data stream, a DQ table is created and named according to the included measurements. The streaming attributes are written in the **Column**. The starting point **T Begin** identifies the corresponding data quality window including **Accuracy** and **Completeness** that are presented as generic Quality Metrics in Figure 27, below.

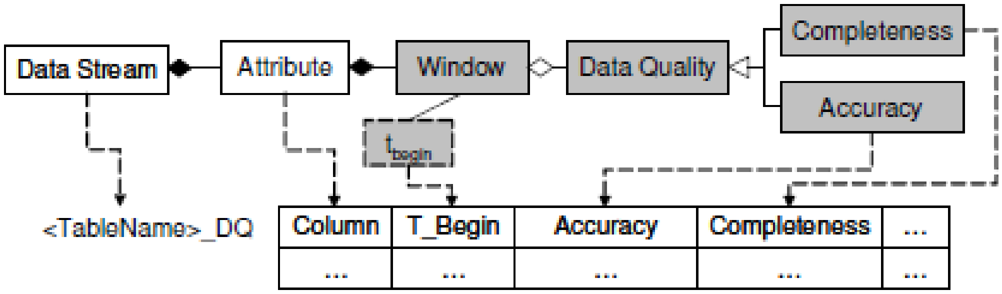


Figure 27: Metadata Mapping

Figure 28 shows what the Jumping Window Meta Mapping looks like for the data processed by Qbase simulator and stored in the corresponding relational database. The Data Quality Table stores aggregated quality metrics for every *n* video frames starting with **FrameID_begin** to **FrameID_end**. The aggregation algorithm is also described in the DQ table.

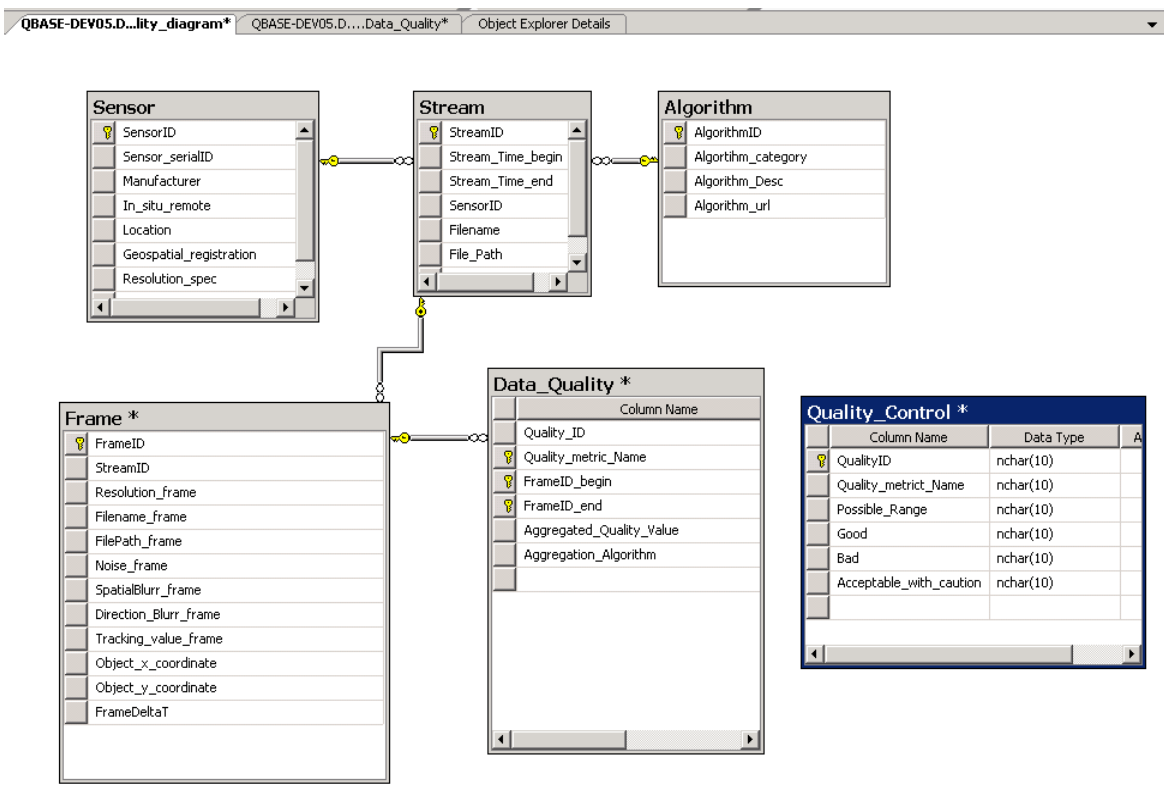


Figure 28: Jumping Window – Data Quality Table Mapping

To Relational Database Model DBMS for CLIF 2006 Ground Camera Data Processed by Qbase Simulator

Figure 29 shows an example of window size aggregated quality metrics for the video data stream:

	Quality_ID	Quality_metric_Name	FrameID_begin	FrameID_end	Aggregated_Quality_Value	Aggregation_Algorithm
1	1	Completeness	1	10	0.75	Linear Average
2	2	Accuracy	1	10	2.35	Linear Average
3	3	Tracking_Consistency	1	10	0.85	Linear Average
4	1	Completeness	11	21	0.95	Linear Average
5	2	Accruacy	11	21	3.1	Linear Average
6	3	Tracking_consistency	11	21	0.9	LInear Average
7	10	Data_integrity	1	4545	1	Random
8	11	Timeliness	1	4545	1	Random
9	25	Aggregated_quality_score	1	4545	0.85	Weighted Average

Figure 29: Window Size Aggregated Quality Metrics Table

Based on CLIF 2006 Video Dataset

NOTE: *The concepts described above were partially implemented within the PSSA.*

3.15 Integration with Persistent Sensor Storage Architecture (PSSA)

The PSSA was developed under the auspices of *Task Order 005*: Persistent Surveillance Data Processing, Storage and Retrieval. The goal of PSSA is to provide a high-performance, flexible infrastructure to support the ingestion, exploitation, integration, storage, and dissemination of data generated by any type of sensor. To accomplish this goal, Qbase developed an architecture based upon the Event Collaboration design pattern.[4]

To communicate sensor data and information derived from the sensor data among processing components of the system, this architecture uses a high performance, low-latency messaging system based on ZeroMQ[5] - which we call the "PSSA Cloud." At its core, the architecture defines two types of processing components: publishers and subscribers. A processing component can be a publisher which publishes events, a subscriber which receives events, or both a publisher and a subscriber.

The PSSA defines different types of services based upon these core component types: Ingestion Services, Application Services, Storage Services, and Dissemination Services.

[4] Event Collaboration, Martin Fowler

[5] ØMQ Messaging System

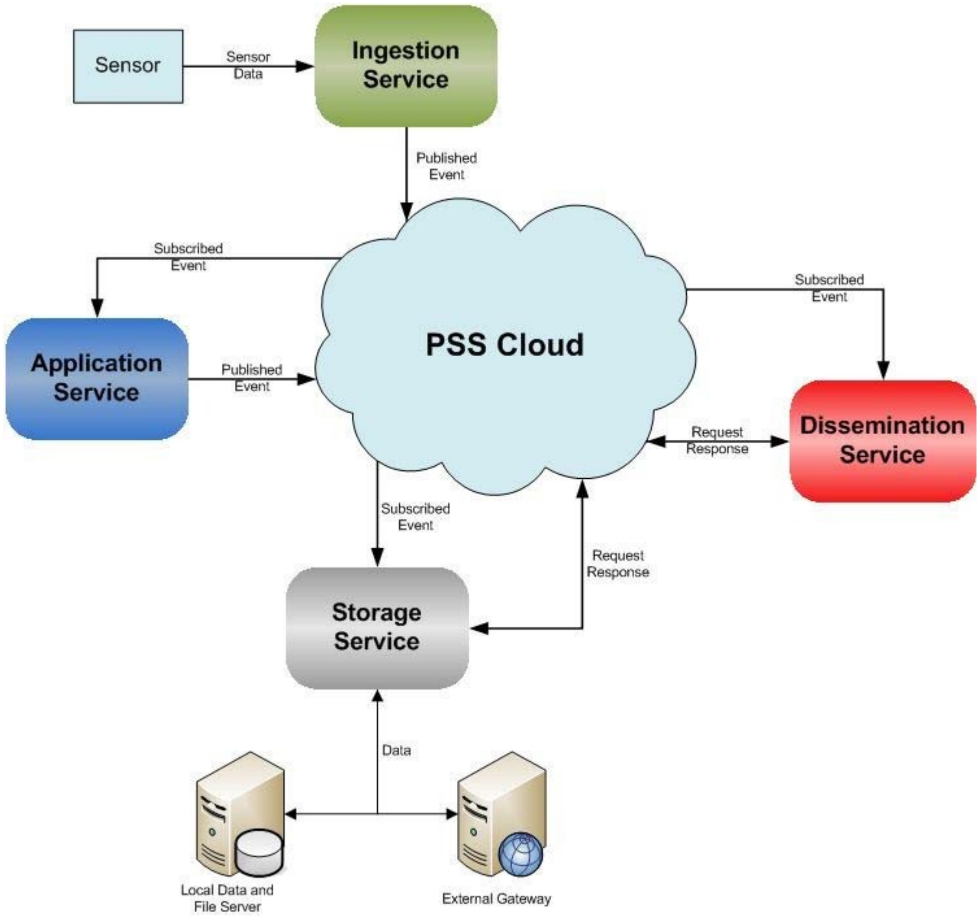


Figure 30: Generic View of Persistent Sensor Storage Architecture

3.15.1. Ingestion Services

Ingestion Services are the processing components responsible for capturing raw sensor data and sensor metadata, formatting and enhancing the data with additional metadata, and then publishing this data as events to the PSSA Cloud. From the perspective of Information Quality, the ingestion component is responsible for creating any quality metadata that is associated with the sensor feed. For example:

- Timeliness of the data – Are we receiving data from the sensor when expected?
- Completeness of the data – Did we receive all of the data expected?
- Integrity of the data – Is the data in the correct format and does it pass basic validation rules?
- Consistency of the data – Does the data make sense based on data received previously? For example, are frame sequence numbers in right order? Are location and time metadata, if present, consistent with the velocity of the sensor platform? And so on.

***NOTE**: The Ingestion Service component's primary responsibility is to get the sensor data and associated metadata into the PSSA Cloud as quickly as possible. Therefore, any quality* analysis of the sensor data beyond quick and simple validations should be deferred to downstream Application Service components.

3.15.2. Application Services

Application Services are processing components that are responsible for the analysis and exploitation of the sensor data. These components subscribe to event messages from Ingestion Services and/or other Application Services in order to generate information required for specific applications. The information generated by the Application Services is in turn published as events for other components of the PSSA system to consume. Examples of Application Services might include object detection and tracking, data enhancement and normalization, geo- registration of the sensor data, and so on. From an Information Quality perspective, the Application Services could be developed to perform quality analysis of sensor data or to aggregate quality metrics from other Application Service components. Most exploitation algorithms implemented by an Application Service will have some type of quality metric associated with it; for example:

- Geo-registration to reference imagery – How well did the points correlate between the sensor data and the reference data?
- Object detection – What is the level of certainty that the object identified is really an object of interest?
- Object tracking – What is the level of certainty that the object being tracked is the same object in subsequent frames?

The Application Services components are the “heavy-lifters” of the system. The PSSA allows these components to be run in parallel with one another on the same or different systems. It allows processing flows to be composed using the event collaboration model.

For example, the Object Tracking Service could use events generated by the Object Detection Service and the Geo-Registration Service, both of which are running independently and know nothing of the Object Tracking Service. Similarly, an Information Quality Application Service could fuse quality metadata from the Sensor Ingestion Service, the detection service, the Geo-Registration Service, and the Tracking Service to determine the reliability of the track information before it’s presented to the user.

3.15.3. Storage Services

Storage Service components are responsible for persisting any data published to the PSSA cloud that needs to be stored, as well as for providing access to that data - and in some cases, externally stored data - to the Application Service and Dissemination Service components.

For the initial implementation of the PSSA reference system, two storage components were developed, one to store streaming media data and another to store all the other data published to the PSSA cloud. Information quality metadata generated by the system is stored by the latter.

The Storage Service components are subscribers to the events generated by other components of the system and typically do not publish events themselves.

Special “gateway” Storage Services can be built to store and retrieve data in systems external to the PSSA cloud. These services may be used by Application Services to get data required to perform their processing or by Dissemination Services to supplement the data stored internally to the system. They are also one means of integrating the PSSA system with external systems such as the DoD Distributed Common Ground Station (DCGS).

3.15.4. Dissemination Services

The primary role of the Dissemination Service components is to make data stored by the Storage Services and/or published to the PSSA cloud available to systems external to the PSSA cloud. Examples of Dissemination Services include a Streaming Media Service used by off the shelf media clients to display real time or stored streaming media and a Web Feature Service (WFS) used to support geospatial queries of sensor metadata. Dissemination Services can also be developed to directly display data published to the PSSA cloud or retrieved from internal or external data sources using storage service components. The PSSA Dashboard and OpenLST clients are examples of these types of Dissemination Services.

3.15.4. Example

As part of the implementation of the PSSA reference system, we developed an Application Service to perform real time video quality estimation. The components of this demonstration system included the components shown in Figure 31, below.

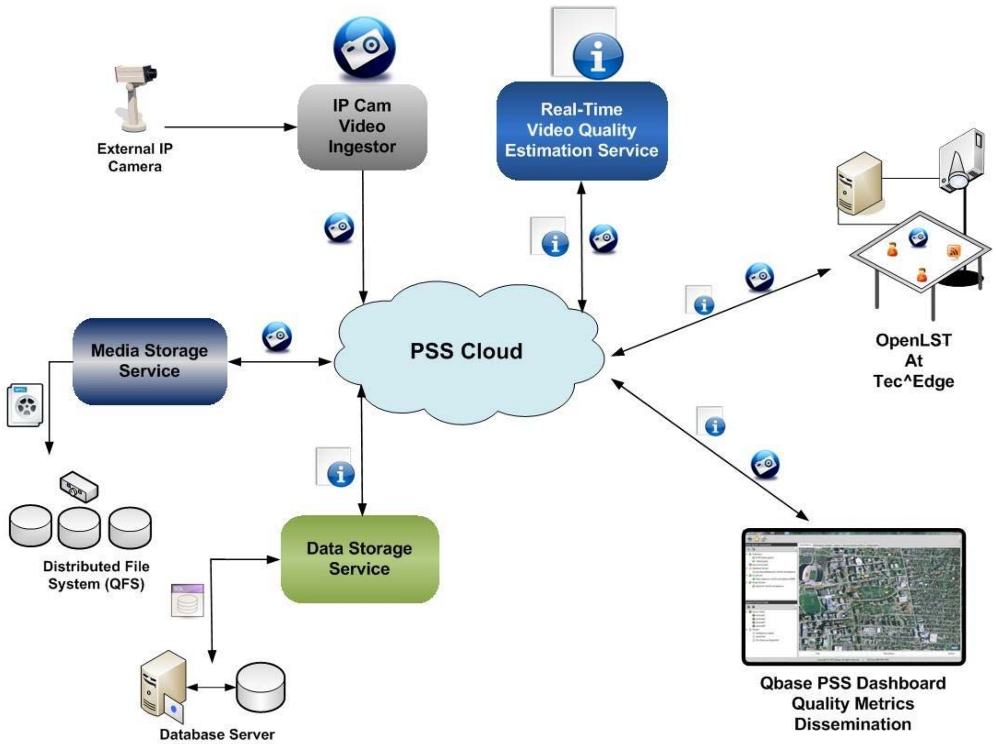


Figure 31: Integration of Video Quality Processing with PSSA

X In this demonstration system, the Real Time Video Quality Estimation Service subscribed to events generated by the Internet Protocol (IP) Camera Video Ingestor which was connected to a live IP video surveillance camera. The Real Time Video Quality Estimation Service sampled the video frames being published and generated data quality scores for the Noise and SSIM quality metrics. These scores were then published as events to the PSSA cloud and picked up by the Data Storage Service to be persisted in the database as well as by the Qbase PSS Dashboard to be displayed alongside the video. The demonstration system also provided a Media Storage Service to record the live video feed and the OpenLST as an alternative to the dashboard for visualizing the live video data.

3.16 Information Quality Scenario

3.16.1. Background

The DoD is increasing the use of network-centric warfare in an attempt to increase mission effectiveness through information sharing and collaboration using distributed battlefield networks. Largely due to developments in technology, the warfighter must manage larger volumes, new forms, and aggregations of complex information than ever before. Interacting with this complex environment that generates, stores, manipulates, accesses, and utilizes an ever-expanding array of electronic resources requires new ways of interacting with information.

In general, the warfighter and their support systems need to be able to:

- Perceive – Recognize relevant information, e.g. sensors, Signals Intelligence (SIGINT), Human Intelligence (HUMINT), etc.
- Comprehend – Process the perceived information in an appropriate way, and
- Project – Synthesize the results into a relevant situational response.

Having access to accurate and timely information is critical to effectively perform mission planning and execute operations. Warfighters need to answer the following questions:

- "How good is the information?"
- "How relevant is the information that I'm being shown?"
- "Do I need to react?"

The PSSA allows quality information to flow with the data throughout the system. For each processing step within the system, the quality information can be examined to aid in deciding the relevancy of the data. For example, an algorithm may decide that a sensor doesn't provide sufficient resolution to provide meaningful results and therefore may decline to process it (e.g. tracking individuals). Although, a human reviewing the output of that same sensor may decide that it's good enough for their purposes (e.g. distinguishing between roads and buildings). Information Quality metrics can assist all users of the data by helping them make a more informed choice regarding the suitability of any given sensor stream to their purposes.

3.16.2. Scenario

Let us suppose that a mission is being planned to assault a compound, suspected to contain a high value target. As mission planning progresses, various data sources are fused together to form a plan of action. These might include low resolution overhead imagery, human intelligence, and other signal intercepts. Using these sources of information, it was determined that the wall surrounding the compound was approximately three meters high and that the target was likely to be on the ground floor of the main building. Ingress and egress routes were planned based on the information at hand.

Additionally, an informant had reported the presence of dogs within the compound, but it was unclear whether they were pets or used as guard dogs. A review of the low resolution imagery provided no indication of the type of trails left by guard dogs; therefore, it was concluded that they were most likely pets.

3.16.3. Example

As the scenario unfolds, additional information becomes available and must be assessed in real-time to ensure the success of the mission. As the strike team is on their way to the target, a UAV is tasked to perform a low level over flight of the compound to provide high resolution imagery that might confirm the presence of the target.

As the video from this UAV over flight is ingested into the PSSA; it is concurrently stored, analyzed and broadcast in real-time to operation controllers. Quality meta-data indicate that although the video is better than that which was used for mission planning; it is of insufficient quality for the human identification algorithms. Therefore, those algorithms do not analyze the data stream and human reviewers are unable to definitively determine if the target is there. However, an algorithm that detects the presence of trails left by guard dogs is able to run and detects the probable presence of a trail that runs just inside the perimeter wall. Based upon the quality meta-data from the sensor source, this detection is considered to be of high quality and is relayed to the strike team to take appropriate measures to deal with this new threat.

As we can see from the above scenario, the addition of quality meta-data with data streams allows both humans and machines to make more informed decisions about the usefulness of any given data set.

4.0 RESULTS AND DISCUSSION

4.1 Phase I

Phase I of Information Quality Tools for Persistent Surveillance Data Sets was primarily dedicated to research and understanding of the current status of Quality of Information in sensor data streams. We studied modern technologies such as SensorML and UncertML which have the potential to incorporate, propagate, and store Quality metrics for sensor data streams along with data stream itself.

We have developed a flexible framework - the Jumping Window/Sampling architecture — in order to monitor quality of data in real time. We have developed an AQS methodology that is based on the statistical analysis and historic trending which can be applied to monitor the quality of information in real time, as well as in forensic mode.

We collected data from different sources and processed the data with various applications in order to come up with a better understanding of the quality metrics in video streams that will contribute into the AQS and give enough confidence and trust into the Quality of Data. In addition, we have incorporated several quality metrics into PSSA and run preliminary sampling calculations of Noise and SSIM metrics with the real time video stream. Both metrics were displayed in the PSSA Dashboard that represents PSSA Dissemination Service.

4.2 Phase II

During Phase II of the Information Quality Tools for Persistent Surveillance Data Sets, we expanded on the work performed during the first year by implementing a schema for communicating and storing information quality metrics in a standardized format and by applying the aggregated quality score methodology to real time and previously recorded sensor data sets. In addition, we developed a model for calculating a metric that utilizes objective and subjective quality information to establish the value of the information for a specific mission. At the end of the second phase we are able to simulate real time data streams using recorded sensor data sets from multiple sensors being ingested into a PSSA reference system.

Once the data is ingested into the PSSA reference system, we are able to simulate the exploitation of this data for the generation of information quality metrics including the value of the information, the storage and retrieval of these metrics, and the visualization of these metrics in conjunction with the sensor data.

Using the simulator, we are able to vary the quality of the sensor data and metadata prior to ingestion into the system, so that we can demonstrate the effects of these variations in the AQS and the resulting value of that information for a specific purpose.

To accomplish these goals, we performed the following tasks:

- Enhanced the Simulator developed for Phase I to read additional sensor data sets and to support the generation/modification of sensor metadata.
- Developed a wrapper for the Persistent Sensor Storage Software Development Kit (PSS/SDK) to allow exploitation algorithms developed in MATLAB by UALR and others to be easily implemented as PSSA Application Services.

- Fully developed an Application Service to implement an AQS for one or more sensor feeds and/or applications (such as object tracking).
- Experimented with different visualization techniques for displaying information quality data to the end user of the system.

5.0 CONCLUSIONS

As stated in the introductory sections of this document, it is critical for data analysts and decision makers to understand the quality of the data upon which a decision to take some action is based. Decisions are based upon the analyst/decision maker's awareness of the situation. At each step within the process of assessing the situation, it is critical to evaluate and communicate the quality of the information that is being generated.

The first step in this process is to measure and provide the ability to communicate the quality of the data being captured through various sensors and sensor systems. We have designed an approach using our PSSA whereby sensor data quality metrics can be calculated and propagated along with the sensor data so that those metrics are available when the data stream is being viewed. These metrics are largely objective in nature and can be determine by performing algorithmic computations to the data. Using the Situational Awareness Model that we described earlier in this document, this corresponds to the sensing stage of assessing the situation.

The next step in this process is to analyze the sensor data to identify the facts about the situation such as what events are taking place and where are they taking place (perception). This stage of assessing the situation can be performed through the use of computer based algorithms or human interaction using annotation tools. Regardless of how this data is generated it is important that the relevant quality measures are included and propagated along with the data. At this stage, there will be objective quality measures based on the quality measures of the source data and subjective quality measures based on how confident the algorithm or human operator is about the result achieved.

It is the responsibility of the algorithm developer or the human operator to determine and communicate this confidence level as part of the analysis and detection process. Signal detection theory provides guiding principles that can be applied to measuring the quality of the results of this processing. This stage of the assessment process is performed in the PSSA using application services that are designed to analyze one or more sensor data streams and detect entities, events, or relationships. The sensor data received by these services is subsequently enhanced with the results of the analysis and associated metrics describing the quality of the information generated. Through the use of a regression model that we described earlier in this document, we propose that an aggregate quality score can be developed to help the analyst understand how all of the quality factors measured up to this stage affect the quality and value of the information presented to them. This information can also be used to "weed out" data that is of such poor quality that it does not make sense to propagate further downstream.

Following analysis of the data to identify the facts of the situation, the next step is to attempt to determine from the facts if there are any activities that might be of interest taking place (comprehension). As with the previous step, this step of the process may be automated or require human interaction or a combination of both. It is important regardless of whether the approach is automated, human based or a combination of the two, that the quality information from previous steps is provided and considered in determining what activities are taking place. Just like in the previous stages, the processing performed during this stage must include metrics that represent the quality of the information being generated. This not only includes the objective quality measures that indicate how accurate, complete, and trustworthy the data is but also more subjective quality measures such as confidence level (i.e. probability of detection, probability of false alarm, etc.), relevance, usefulness.

As described in this document, we propose that these quality measures can be combined using a binary logistic regression model to determine the overall value of this information to the analyst for a given mission objective. One of the goals of this approach is to prioritize the data, so that downstream algorithms and analysts can focus on information that provides the most value to the mission objectives.

Using the data collected and information generated in the previous phases (including the quality and value of information metrics), the next step is to determine whether the activity or activities detected indicates the potential that an undesirable situation exists or will develop that requires some type of action to be taken (projection).

Automated processes may be used to identify the existence of or potential for threatening situations to develop, however, before any action is taken, a human must review and confirm the results of the automated process. It is critical that this person have visibility into the quality of the data used to project the potential outcome(s). Just as in the previous steps, objective and subjective quality measures should be used to determine how much trust can be placed in the machine results as well as how likely it is that a projected outcome will occur. The individual objective and subjective quality metrics captured throughout the situation assessment process all contribute to the overall level of trust in the information and data presented to the decision maker. The decision maker must take these factors into account in determining whether to take action. Therefore it is critical that the quality and value of the information be captured and propagated to the decision maker along with the information, itself.

6.0 RECOMMENDATIONS

Our recommendation for future work in this area focuses on practical application of the approaches and principles outlined in this document. During the first two phases of this project, we have developed a platform based on the persistent sensor storage architecture and using the sensor simulator that allows us to use previously recorded sensor data to measure the impact of degrading various aspects of the sensor data (reducing resolution, decreasing frame rate, introducing compression artifacts, adding noise, etc.) on both image processing algorithms and human perception.

Over the upcoming year, we plan to refine the Value of Information model described in this document through a series of experiments using the existing PSSA platform and the sensor simulator. We anticipate that minor enhancements will be needed to both in order to adapt them to the type of sensor data and exploitation algorithms that are available and which provide the scenarios that we want to investigate.

The core focus of the next phase of this project is to identify and collect sensor data for a variety of different scenarios with different mission objectives that are relevant to military and civilian persistent surveillance applications. This data will be used to run experiments that include varying the quality and value of information provided to subjects attempting to achieve the mission objectives. Using subjective measures provided by the test subjects we will attempt to build models that predict how the quality and value of information parameters affect the ability of those subjects to accomplish their mission objectives.

We will then test these measures against different scenarios that have the same mission objectives to determine whether the objective and subjective quality measures determined by analyzing and processing the data can be used to predict the value of that information in accomplishing the mission. Our intention is to use this approach to predict the value of information generated during both the comprehension and the projection stages of situational assessment.

7.0 REFERENCES

[1] A. Moorthy, A. Bovik, "Visual Importance Pooling for Image Quality Assessment, *IEEE Journal of Selected Topics in Signal Processing*, Vol 3, No 2, April, 2009.

[2] S. Thompson, J. Kastanowski, S. Fairgrieve, "PULSENetTM, " Military Communications Conference, 2006, MILCOM 2006, IEEE.

[3] L. Bastin, D. Cornford, M. Williams, "Computer Says Maybe…," Aston University PPT presentation.

[4] M. Williams, D. Cornford, L. Bastin, E. Pebesma, "Uncertainty Markup Language (UncertML), " Discussion Paper, Document Number: 08-122r1.

[5] E. Blasch, S. Plano, "JDL Level 5 Fusion Model "User Refinement, " *Issues and Applications in Group Tracking*, SPIE Vol 4729, Aerosense, 2002, pp. 270 – 279.

[6] M. Anwar Hossain, Pradeep K. Atrey, and Abdulmotalrb El Saddik, "Modeling Quality of Information in Multi-Sensor Surveillance Systems," *ICDE*, 2007.

[7] P. L. Callet, F. Autrusseau, "Subjective Quality Assessment IRCCyN/IVC Database," http://www.irccyn.ec-nantes.fr/ivcdb/

[8] H.R. Sheikh, Z. Wang, L. Cormack, and A.C. Bovik, "LIVE Image Quality Assessment Database Release http://live.ece.utexas.edu/research/quality

[9] E. Mendi, M. Milanova, Y. Zhou, J. Talburt – University of Arkansas at Little Rock. Presented at the 9th WSEAS International Conference on Signal Processing, Robotics, and Automation; Cambridge, UK; February 20-22, 2010.

[10] http://www.compression.ru/video/

[11] Chatschik Bisdikian(a), Lance M. Kaplan(b), Mani B. Srivastava(c), David J. Thornley(d), Dinesh Verma(a), and Robert I. Young(e) "Building Principles for a Quality of Information Specification for Sensor-originated Information," IBM Research Report: RC24762, Computer Science, March17, 2009.

[12] VIRAT Video Dataset Scoring Software Ver. 1.2 http://www.viratdata.org/

[13] VIRAT_Video_Dataset_Release1.0_Introduction_v1.0.pdf http://www.viratdata.org/

[14] "A Large-Scale Benchmark Dataset for Event Recognition in Surveillance Video" by Sangmin Oh, Anthony Hoogs, Amitha Perera, Naresh Cuntoor, Chia-Chih Chen, Jong Taek Lee, Saurajit Mukherjee, J.K. Aggarwal, Hyungtae Lee, Larry Davis, Eran Swears, Xiaoyang Wang, Qiang Ji, Kishore Reddy, Mubarak Shah, Carl Vondrick, Hamed Pirsiavash, Deva Ramanan, Jenny Yuen, Antonio Torralba, Bi Song, Anesco Fong, Amit Roy-Chowdhury, and Mita Desai, in Proceedings of IEEE Comptuer Vision and Pattern Recognition (CVPR), 2011.

[15] Anja Klein, Hong Hai Do, Gregor Hackenbroich, Marcel Karnstedt, Wolfgang Lehner, “Representing Data Quality for Streaming and Static Data, ” International Conference on Data Engineering – ICDE.

[16] J. Talburt, S. Dagtas, M. Milanova, M. Tudoreanu, “Information Quality Tools for Persistent Surveillance Data Sets, *UALR*, June 2010

[17] Z. Wang and A. C. Bovik, “Mean Squared Error: Love It or Leave It?”, *IEEE Signal* Processing Magazine, January 2009.

[18] Z. Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli, “Image Quality Assessment: from Error Visibility to Structural Similarity, *IEEE Transactions on Image Processing*, Vol.13, No.4, pp. 600- 612, April 2004.

[19] H. R. Sheikh, M. Sabir, and A. C. Bovik, “A Statistical Evaluation of Recent Full Reference Image Quality Assessment Algorithms,” *IEEE Trans. Image Process*., Vol. 15, No. 11, pp. 3440–3451, Nov. 2006.

[20] H.R. Sheikh and A.C. Bovik, “Image information and visual quality,” *IEEE Transactions on Image Processing*, Vol.15, No.2,pp. 430- 444, Feb. 2006.

[21] M. Milanova , S. Rubin, R. Kountchev, V. Todorov, and R. Kountcheva, “Combined Visual Attention Model for Video Sequences,”*IEEE ICPR* 2008, 1-4.

[22] K. Seshadrinathan, R. Soundararajan, A. C. Bovik, and L. K. Cormack, “Study of Subjective and Objective Quality Assessment of Video, ” *IEEE Transactions on Image Processing*, Vol.19, No.6, pp.1427-1441, June 2010.

[23] K. Seshadrinathan, R. Soundararajan, A. C. Bovik, and L. K. Cormack, “A Subjective Study to Evaluate Video Quality Assessment Algorithms, ” *SPIE Proceedings Human Vision and Electronic Imaging*, Jan. 2010.

[24] Mica R. Endsley. “Toward a Theory of Situation Awareness in Dynamic Systems,” *Human Factors Journal*, Vol. 37(1), pp 32-64, March 1995.

[25] J. Salerno, M. Hinman, D. Boulware, “Building A Framework For Situation Awareness, ” Proceedings of the Seventh International Conference on Information Fusion, Vol. I, pp. 219-226, 2004

[26] Barry McGuinness and Louise Foy, “A Subjective Measure of SA: The Crew Awareness Rating Scale (CARS),” *Proceedings of the First Human Performance, Situation Awareness, and Automation Conference*, Savannah, Georgia, October 2000.

[27] Abdi, H. (in press, 2009). Signal detection theory. In McGaw, B., Peterson, P.L., Baker, E. (Eds.): *Encyclopedia of Education (3rd Ed).* New York: Elsevier.

APPENDIX A - Objective Metrics Implemented

These metrics were described in UALR Final Report [16]. Here we just mention a few of them.

Mean Square Error (MSE): MSE is widely used as it is parameter free, computationally simple and mathematically convenient in the context of optimization. It also represents image energy measure that energy is preserved after any orthogonal linear transformation, such as the Fourier transform. However, MSE does not fit precisely with the perceived visual quality. Distorted images with the same MSE may have different visibility [17], [18].

Consider two images $x = \{x_i \mid i = 1, 2, \ldots, N\}$ and $y = \{y_i \mid i = 1, 2, \ldots, N\}$ where N is the number of pixels and x_i and y_i are the i th pixels of the images of x and y, respectively; the MSE between these two images is:

$$\mathrm{MSE}(x, y) = \frac{1}{N} \sum_{i=1}^{N} (x_i - y_i)^2 \qquad (7)$$

Structural SIMilarity Index (SSIM): Consider two images $x = \{x_i \mid i = 1, 2, \ldots, N\}$ and $y = \{y_i \mid i = 1, 2, \ldots, N\}$ where N is the number of pixels and x_i and y_i are the i th pixels of the images of x and y, respectively. SSIM- SSIM (x, y) combines three comparison components, namely luminance- l (x, y), contrast-c(x, y) structure- s(x, y) [19]:

$$\mathrm{SSIM}(x, y) = f(l(x, y), c(x, y), s(x, y)) \qquad (8)$$

Luminance, contrast and structure comparisons are defined as follows:

$$l(x, y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1}, \quad C_1 = (K_1 L)^2$$

$$c(x, y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2}, \quad C_2 = (K_2 L)^2 \qquad (9)$$

$$s(x, y) = \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3}, \quad C_3 = \frac{C_2}{2}$$

Where:

μ_x, μ_y, σ_x, σ_y and σ_{xy} are means of x and y, variances of x and y and correlation coefficient between x and y. K_1 and K_2 are scalar constants that $K_1, K_2 << 1$ and L is the dynamic range of the pixel values. Finally, SSIM index yields to:

$$\mathrm{SSIM}(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \tag{10}$$

Weighted Objective Quality Metric When the Task is Tracing Moving Objects in Video: In human visual system, the importance of a visual event should increase with the information content, and decrease with the perceptual uncertainty [20], we incorporated foreground mask as weighting function into the MSE and SSIM metrics to measure the motion feature of the moving car. At a time MSE is MSE (x, y, t)and SSIM is SSIM (x, y, t). The weighting function is:

$$w(x,y,t) = \left| I(x,y,t) - median\ \{I(x,y,t-i\} \right| > \tau \tag{11}$$

We define weighted MSE as wMSE and weighted SSIM as wSSIM as follows:

$$\begin{aligned} w\mathrm{MSE} &= \frac{\sum_x \sum_y \sum_t w(x,y,t)\mathrm{MSE}(x,y,t)}{\sum_x \sum_y \sum_t w(x,y,t)} \\ w\mathrm{SSIM} &= \frac{\sum_x \sum_y \sum_t w(x,y,t)\mathrm{SSIM}(x,y,t)}{\sum_x \sum_y \sum_t w(x,y,t)} \end{aligned} \tag{12}$$

APPENDIX B - Spatial/Temporal Quality Metadata

***Spatial Information Quality Metadata*:** The metadata used to determine the initial coverage area of the sensor should be evaluated to determine the accuracy of that coverage area. For 2-dimensional (2D) locations, the circular error associated with the location data should be determined and included as part of the information quality metadata.

For 3-dimensional (3D) locations, the spherical error should be determined and included as well. This will allow the accuracy of the location information to be normalized and reported across multiple sensors regardless of the source of the location information. The sensor metadata used to determine the circular error and/or spherical error should be reported as well.

For example, if Global Position System (GPS) data is used to locate the sensor, each GPS reading should have the Dilution of Precision (DOP) data as part of its metadata. DOP is typically expressed in two forms: Horizontal Dilution Of Precision (HDOP) for latitude/longitude precision and Positional Dilution Of Precision (PDOP) for latitude/longitude/altitude precision. However, if the sensor does not provide this information, a theoretical DOP for any given time and location can be calculated using a GPS Satellite Almanac (ephemeris data) and assumptions regarding which satellites are visible to the receiver.

For every sensor in the PSSA system that includes location metadata, the accuracy of its location measuring device (for example, GPS) should be identified as part of the sensor metadata. Positional accuracy for GPS devices is typically based on the probability that the reading provided by the device falls within a circle whose radius is the accuracy value and whose center is the actual location. Figure B-1 shows an example of this with two probabilities (50% and 95%).

Historically, the military has used Circular Error Probable (CEP) for specifying location error. CEP is a 50 percentile circular distribution - meaning that at least 50% of the location readings will be within the specified radius of the actual location. Most GPS manufacturers use a 95 percentile value when publishing their accuracy data.

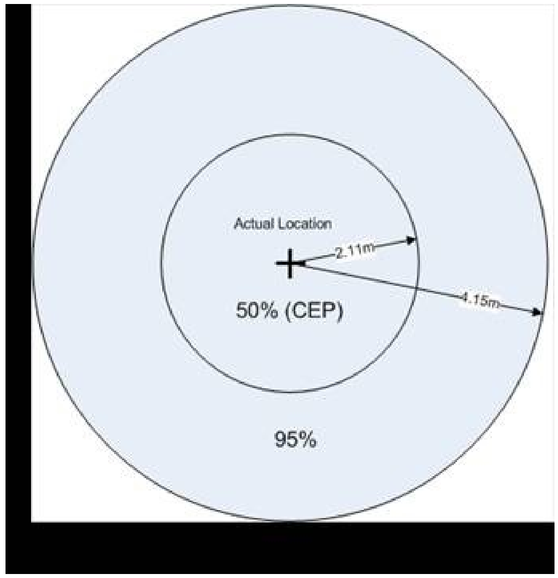


Figure B-1: GPS Accuracy Example

For the example shown in Figure B-1 above, the CEP accuracy value is 2.11m and the 95th percentile accuracy value is 4.15m. This means that at least 50% of the readings provided by the device will be within 2.11m of the actual location and 95% of the readings will be within 4.15m of the actual location. These figures can be combined with the GPS HDOP and/or PDOP values to provide an estimate of the circular error and/or spherical error associated with a specific location reading.

For an excellent overview of GPS accuracy see the following article from the January 2007 issue of GPS World: http://www.gpsworld.com/lbs/infrastructure/gnss-accuracy-lies-damn-lies-and-statistics-1771?page_id=5

As a rule of thumb, the accuracy associated with a GPS reading can be determined by multiplying the published accuracy by the HDOP or PDOP value to produce a circular or spherical error value for the GPS reading. This error value should be included with location metadata. In order to normalize this error value for all location readings, the 95th percentile accuracy value for the GPS device should be used. If the GPS manufacturer uses a different percentile, then it can be converted to the 95th percentile as described by the following:

http://www.gpsworld.com/files/gpsworld/nodes/2007/1771/i9.jpg

The DOP value provided can also be used to determine whether or not to use the GPS data. A general rule of thumb is to take the published accuracy of the GPS device and multiply it by the DOP value to get a maximum error for the GPS reading. For example, if the accuracy of the GPS device is +/- 3m and the DOP value is 3, then the actual location is within 9m of the GPS reading (3m x DOP of 3 = 9m).

For reference, the table below provides interpretations of the DOP values from two different Internet sources:

Table B-1: DOP Values from Two Different Internet Sources

DOP Value[6]	DOP Value[7]	Rating	Description
1	1	Ideal	This is the highest possible confidence level to be used for applications demanding the highest possible precision at all times.
2-3	1-2	Excellent	At this confidence level, positional measurements are considered accurate enough to meet all but the most sensitive applications.
4-6	2-5	Good	Represents a level that marks the minimum appropriate for making business decisions. Positional measurements could be used to make reliable in-route navigation suggestions to the user.
7-8	5-10	Moderate	Positional measurements could be used for calculations, but the fix quality could still be improved. A more open view of the sky is recommended.
9-20	10-20	Fair	Represents a low confidence level. Positional measurements should be discarded or used only to indicate a very rough estimate of the current location.
21-50	>20	Poor	At this level, measurements are inaccurate by as much as 300 meters with a 6 meter accurate device (50 DOP × 6 meters) and should be discarded.

These should be viewed as guidelines since the accuracy level of GPS devices vary (for example, if a GPS device has a 95th percentile accuracy of 6m, than even a DOP of 1 will only ensure accuracy to within 6m). Some GPS devices support Differential GPS (DGPS) and/or the Wide Area Augmentation System (WAAS) which can significantly increase the accuracy of the GPS reading. The accuracy metadata for the GPS reading should reflect the improved accuracy if DGPS or WAAS is used.

For spatial measurements, we are primarily concerned with the PDOP and the HDOP. HDOP represents the dilution of precision in 2D space (latitude/longitude) and PDOP represents the dilution of precision in 3D space (latitude/longitude/altitude). HDOP and PDOP can be used to estimate the circular error and spherical error, respectively of the GPS location. The circular error represents the error in 2D and is calculated by multiplying the accuracy of the GPS sensor by the HDOP value. For example, a GPS device with an accuracy of 6m and a HDOP of 1.5 will yield a circular error of 9m. Similarly, the same device with a PDOP of 2 will have a spherical error of 12m.

[6] Source 1: http://www.geoframeworks.com/articles/WritingApps2_3.aspx
[7] Source 2: http://en.wikipedia.org/wiki/Dilution_of_precision_(GPS)#Meaning_of_DOP_Values

The Information Quality Metadata associated with a GPS location should include the dilution of precision data provided by the sensor or calculated from a satellite almanac, the source of the DOP data (sensor vs. almanac), and an estimation of the circular error in meters based on the DOP data. For 3D data, the metadata should also include an estimate of the spherical error associated with the location. For locating devices other than GPS, a circular error in meters should still be provided based on the particular characteristics of the locating device.

Temporal Information Quality Metadata: For each sensor within the system, we need to capture the precision and accuracy of the time source used by the sensor, if it has one, as well as the precision and accuracy of the time source used by the ingestion components. The precision of the time source is typically a fixed value based on the resolution of the sensor's time reporting mechanism and does not change from reading to reading. The time precision of the sensor should be recorded within the system as part of the sensor's metadata.

On the other hand, the accuracy of the time source could change from reading to reading. Many sensors rely on GPS receivers to provide their time context, the amount of error in the time reported by the GPS receiver is related to the Time Dilution of Precision (TDOP) of the GPS reading. Most GPS receivers do not provide this information as part of their metadata stream. However, if the Satellite Vehicles (SVs) that were used to determine the time are known, the TDOP can be computed using ephemeris data from a GPS Satellite Almanac.

In order to accommodate variations in the precision of the presentation time between different sensor types, this value is actually stored as a time span (start-time/end-time) during which the ingestion component is *xx*% confident that the sensor reading occurred. This level of confidence should be captured as additional information quality metadata. For reporting/displaying presentation time, the midpoint between start time and end time is used.

For each sensor, there is some time-related Information Quality Metadata which should be captured and passed along with the time information. The time metadata associated with each sensor reading includes all of the times listed above along with the information quality metadata and statistics listed below:

- Accuracy/Reliability of the Acquisition Time Data — this would typically be reported by the sensor as part of its metadata stream. If acquisition time is not provided by the sensor then this metadata will not be present.
- Accuracy/Reliability of the Ingestion Time Data — this information is determined by the ingestion component used to bring the sensor data into the system.
- Accuracy/Reliability (Confidence) of the Presentation Time Data — this information is determined by the ingestion component based upon whatever algorithm/conversions are used to determine the presentation time.
- Latency statistics:
    - Delta between acquisition time and ingestion time, if known
    - Average delta (moving average)

- Deviation of current delta from moving average
- Deviation of current delta from expected delta

- Acquisition time statistics:
  - Delta between current acquisition time and previous acquisition time
  - Average delta (moving average)
  - Deviation of current delta from moving average
  - Deviation of current delta from expected delta
- Ingestion time statistics:
  - Delta between current ingestion time and previous ingestion time
  - Average delta (moving average)
  - Deviation of current delta from moving average
  - Deviation of current delta from expected delta

The ingestion component is responsible for tracking this quality metadata and also providing a Presentation time span for which the sensor data can be considered valid. This should be a time interval for which we are *xx*% confident that the sensor reading was captured. A confidence metric is reported as part of the information quality metadata associated with the presentation time that reflects the accuracy/reliability of the presentation time span.

APPENDIX C - Jumping Window Detailed Description

A sensor data stream *D* of length *m* and rate *r* consists of *n+1* Attributes Ai (0 <=*i* <=*n*), where A^0 represents the timestamp *t* of the sensor data stream. Each timestamp t_j (0<= *j* <= *m*) indicates a tuple T_j with *n* measurement values v_{ij} .

Every measurement value v_{ij} is enhanced by the data quality information; for instance, as it is shown in Figure C-1, accuracy and completeness. Obviously, this approach significantly increases the data volume, which is multiplied by the number of considered DQ dimensions. So, to reduce the volume of metadata while preserving the concept of enhancing the data stream with quality metadata, we introduce the Jumping Window architecture.

Each measurement attribute stream is divided into an unlimited number of windows with a given size *s* containing sensor data items (white) and data quality information (gray). Each window is identified by its starting point t_{begin} and consists of *s* measurement values v_{ij}($k <= j <= k + s - 1$) of a certain attribute A_i. Furthermore, the window contains one value for each DQ dimension q_{ik} (for example, window completeness c_{ik} and window accuracy a_{ik}). The number of data quality dimensions is not fixed but can vary for each attribute. The window size *s* can be defined independently for each stream attribute.

For Jumping Window-based annotations, the data quality information is not sent together with every single data item, but rather window-wise for each DQ dimension. The additional data volume is reduced to an acceptable degree by aggregating the data quality for each attribute A_i in jumping stream windows w_{ik} of the given size s_i starting at timestamp t_{begin}. This prevents the real time data stream and the storage from being overloaded. Aggregation functions can be flexibly determined for each DQ dimension depending on the application.

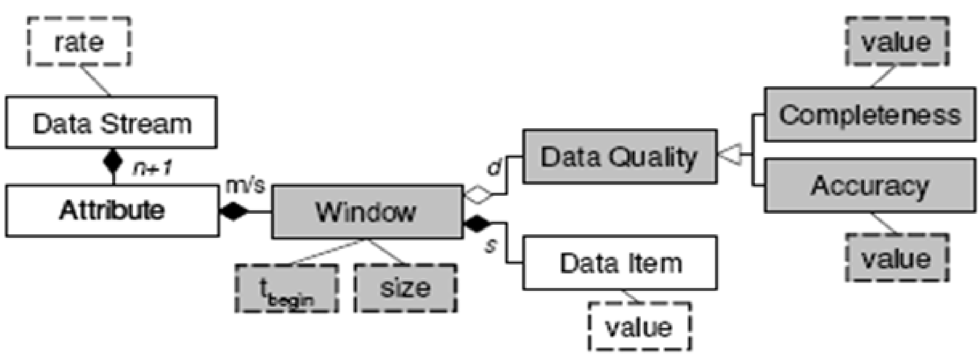


Figure C-1: Jumping Window Architecture for the Propagation of Data Quality (DQ)[15].

The Jumping Window concept was implemented using the Qbase Simulator. Quality metrics were processed and aggregated for every three video frames (see Figure C-2below).

Timestamp	300	600	900	1200	1500	1800	2100	2400	2700	3000
Filename(frame) (image as a measure)										
Video Quality										
Noise	4.1	2.8	6.2	8.4	5.4	5.6	7.8	3.3	4.8	5.2
Blur	4.5	6.5	7.9	4.9	6.3	8.8	9.6	10.1	9.9	9.2
Blocking	13.9	11.0	10.0	12.1	11.9	8.1	12.6	11.8	11.6	11.0
SSIM	83	88	87	82	85%	77	89	98	99	95%
Completeness	85	90	91	83	87%	68	72	78	78	75%
Tracking Quality										
Car Tracking accuracy	67	75	78	57	69%	69	69	69	71	69%
Car Tracking consistency (???)	45	54	67	36	57%	55	36	61	48	55%
Timeliness	100	100	100	100	100	100	100	100	100	100
Usefullness	100	100	100	100	100%	100	100	100	100	100%

Figure C-2: Quality Metrics Aggregated for Every Three Video Frames

LIST OF SYMBOLS, ABBREVIATIONS, AND ACRONYMS

2D	Two Dimensional
3D	Three Dimensional
AQS	Aggregated Quality Score
CEP	Circular Error Probable
CLIF	Columbus Large Image Format
CSUAV	Columbus Surrogate Unmanned Aerial Vehicle Data
CVPR	Computer Vision and Pattern Recognition
DBMS	Database Management System
DARPA	Defense Advanced Research Project Agency
DCGS	Distributed Common Ground Station
DGPS	Differential GPS
DMOS	Difference Mean Opinion Scores
DoD	Department of Defense
DOP	Dilution of Precision
DSMS	Data Stream Management System
DQ	Data Quality
FP/NORM	False Positive/Normalizing factor
GI	Geographic Information
GIS	Geographic Information Science
GML	Geography Markup Language
GPS	Global Positioning System
HD	High Definition
HDOP	Horizontal Dilution of Precision
HUMINT	Human Intelligence
IVC	Irvine Valley College
JPEG	Joint Photographic Experts Group
LAIR	Large Area Image Recorder
LAR	Locally Adaptive Resolution
LCS	Location Services Clients
LIVE	Laboratory for Image and Video Engineering (University of Texas at Austin)
LRM	Linear Regression Model
MATLAB	Matrix Laboratory – a numerical computing environment and fourth- generation programming language developed by MathWorks
MOS	Mean Opinion Score
MPEG	Moving Picture Expert Group
MSAD	Mean Absolute Difference
MSE	Mean Square Error
MSU	Moscow State University (of Instrument Engineering and Computer Science)

O&M	Observations & Measurements
OGC	Open Geospatial Consortium
PDOP	Positional Dilution Of Precision
PSNR	Peak Signal-to-Noise Ratio
PSS	Persistent Sensor Storage
PSSA	Persistent Sensor Storage Architecture
PSS/SDK	Persistent Sensor Storage Software Development Kit
PULSENet	Persistent Universal Layered Sensor Exploitation Network
QoI	Quality of Information
SDT	Signal Detection Theory
SensorML	Sensor Model Language, an eXtensible Markup Language (XML)
SOS	Sensor Observation Service
SPS	Sensor Planning Service
SIGINT	Signals Intelligence
SSIM	Structural SIMilarity Index (a method for measuring the similarity between two images)
SV	Satellite Vehicles
SWE	Sensor Web Enablement
TDOP	Time Dilution of Precision
TML	Transducer Markup Language
UALR	University of Arkansas (Little Rock)
UAV	Unmanned Aerial Vehicle
UncertML	Uncertainty Markup Language
URI	Uniform Resource Identifier
VIVID	Video Verification of Identity
VIRAT	Video Image Retrieval and Analysis Tool
VoI	Value of Information
VQEG	Video Quality Experts Group
WAAS	Wide Area Augmentation System
WFS	Web Feature Service
WNS	Web Notification Service
XML	Extensible Markup Languages